Deep learning and generative AI

Creative artificial intelligence

Tabular data

Play tennis or not? (depending on weather)

Features

Labels

Items

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	90	FALSE	No
Sunny	Hot	87	TRUE	No
Overcast	Hot	93	FALSE	Yes
Rainy	Mild	89	FALSE	Yes
Rainy	Cool	79	FALSE	Yes
Rainy	Cool	59	TRUE	No
Overcast	Cool	77	TRUE	Yes
Sunny	Mild	91	FALSE	No
Sunny	Cool	68	TRUE	Yes
Rainy	Mild	80	FALSE	Yes
Sunny	Mild	72	TRUE	Yes
Overcast	Mild	96	TRUE	Yes
Overcast	Hot	74	FALSE	Yes
Rainy	Mild	97	TRUE	No

Play tennis or not? (depending on weather)

Dataset $\mathbb{D} = \{\langle \mathbf{x}_i, y_i \rangle\}_{i=1}^{n} = \{\langle \mathbf{x}_1, y_1 \rangle, \langle \mathbf{x}_2, y_2 \rangle, \ldots, \langle \mathbf{x}_n, y_n \rangle\}$

items

labels

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	90	FALSE	No
Sunny	Hot	87	TRUE	No
Overcast	Hot	93	FALSE	Yes
Rainy	Mild	89	FALSE	Yes
Rainy	Cool	79	FALSE	Yes
Rainy	Cool	59	TRUE	No
Overcast	Cool	77	TRUE	Yes
Sunny	Mild	91	FALSE	No
Sunny	Cool	68	TRUE	Yes
Rainy	Mild	80	FALSE	Yes
Sunny	Mild	72	TRUE	Yes
Overcast	Mild	96	TRUE	Yes
Overcast	Hot	74	FALSE	Yes
Rainy	Mild	97	TRUE	No

$\mathbf{x}_i = \langle \text{Sunny, Hot, 87, True} \rangle$

$y_i = \langle \text{No} \rangle$

X : set of instances (domain set) ⟷ Every possible weather

Y : set of labels ⟷ {YES,NO} (play tennis or not)

$\mathbf{x} \sim \mathscr{D}(X)$ ⟷ The probability distribution of the weather

$f : X \to Y$: ideal classifier ⟷ A classifier that always correctly predicts if the weather is good to play tennis

$H = \{h \mid h : X \to Y\}$: set of predictors $X \to Y$ ⟷ A set of predictors, e.g. the set of all decision trees that check the specific features

$L_{(\mathscr{D},f)}(h) = \Pr_{\mathbf{x} \sim \mathscr{D}(X)}\left(h(x) \neq f(x)\right)$ the true error ⟷ The probability for a specific predictor to incorrectly predict if the weather is good to play tennis

$L_{\mathbb{D}}(h) = \dfrac{\left|i \in \mathbb{N}_n : h(\mathbf{x}_i) \neq y_i\right|}{|n|}$ the empirical error ⟷ How many times the predictor incorrectly predicts if the weather is good to play tennis, for the specific dataset items?

$h_{\mathbb{D}} \in \arg\min_{h \in H}\{L_{\mathbb{D}}(h)\}$ ⟷ A predictor (from the set of predictors) with the minimal wrong predictions for the specific dataset

Machine learning problem statement

Setting

X : set of instances (domain set)

Y : set of labels

$\mathbf{x} \sim \mathscr{D}(X)$

Dataset $\mathbb{D} = \{\langle \mathbf{x}_i, y_i \rangle\}_{i=1}^{n} = \{\langle \mathbf{x}_1, y_1 \rangle, \langle \mathbf{x}_2, y_2 \rangle, \ldots, \langle \mathbf{x}_n, y_n \rangle\}$

$f : X \to Y$: ideal classifier

$H = \{h \mid h : X \to Y\}$: set of predictors $X \to Y$

features

labels

items

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	90	FALSE	No
Sunny	Hot	87	TRUE	No
Overcast	Hot	93	FALSE	Yes
Rainy	Mild	89	FALSE	Yes
Rainy	Cool	79	FALSE	Yes
Rainy	Cool	59	TRUE	No
Overcast	Cool	77	TRUE	Yes
Sunny	Mild	91	FALSE	No
Sunny	Cool	68	TRUE	Yes
Rainy	Mild	80	FALSE	Yes
Sunny	Mild	72	TRUE	Yes
Overcast	Mild	96	TRUE	Yes
Overcast	Hot	74	FALSE	Yes
Rainy	Mild	97	TRUE	No

Problem

Find $h \approx f$, $h \in H$

Usually

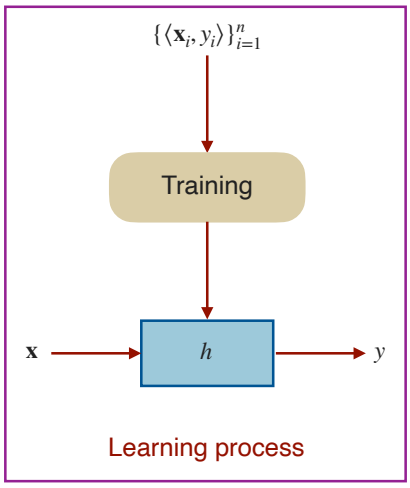
f is unknown

$\mathscr{D}(X)$ is unknown

We assume that $\mathbf{x}_i \sim \mathscr{D}(X)$

$\Rightarrow \mathbb{D}$ is a the only window for observing f and $\mathscr{D}$

Machine learning problem statement



Setting

X : set of instances (domain set)

Y : set of labels

$\mathbf{x} \sim \mathcal{D}(X)$

Dataset $\mathbb{D} = \{\langle \mathbf{x}_i, y_i \rangle\}_{i=1}^n = \{\langle \mathbf{x}_1, y_1 \rangle, \langle \mathbf{x}_2, y_2 \rangle, \ldots, \langle \mathbf{x}_n, y_n \rangle\}$

$f : X \to Y$: ideal classifier

$\mathrm{H} = \{h \mid h : X \to Y\}$: set of predictors $X \to Y$

Problem

Find $h \approx f$, $h \in \mathrm{H}$

$h_{\mathbb{D}} \leftarrow \text{Training}\left(\{\langle \mathbf{x}_i, y_i \rangle\}_{i=1}^n\right)$

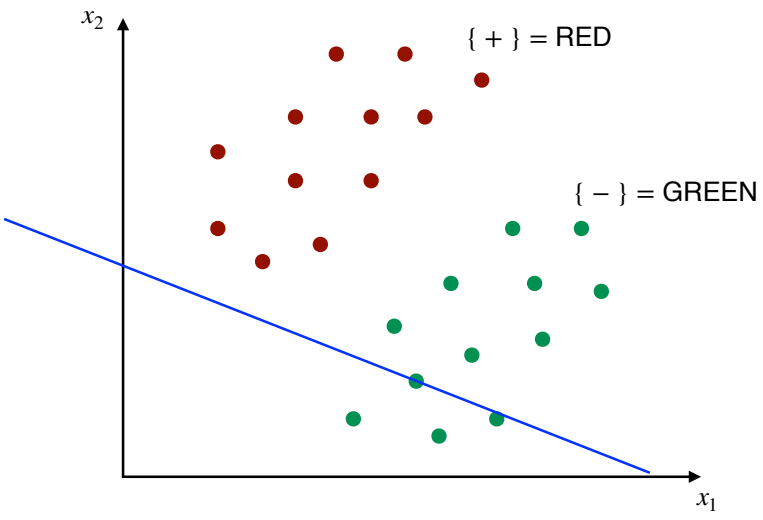
Solving the problem

$h_{\mathbb{D}} \in \arg\min_{h \in \mathrm{H}} \{L_{\mathbb{D}}(h)\}$

Error $L_f(h) = \dfrac{\left|x \in X : h(x) \neq f(x)\right|}{|X|}$ the true error

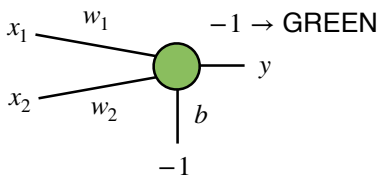
$L_{(\mathcal{D},f)}(h) = \Pr_{\mathbf{x} \sim \mathcal{D}(X)}\left(h(x) \neq f(x)\right)$

$L_{\mathbb{D}}(h) = \dfrac{\left|i \in \mathbb{N}_n : h(\mathbf{x}_i) \neq y_i\right|}{|n|}$ the empirical error



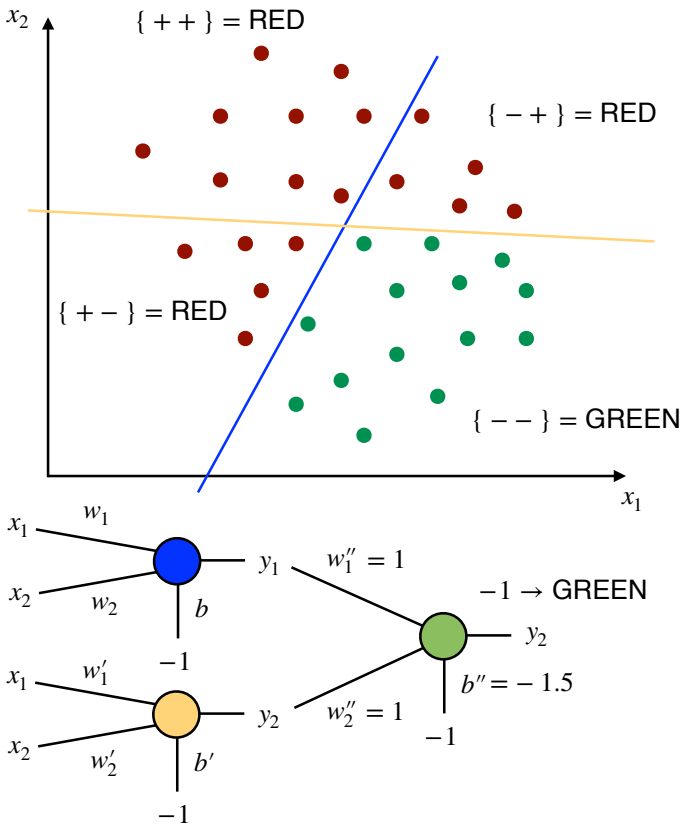
Linear predictors

Perceptron

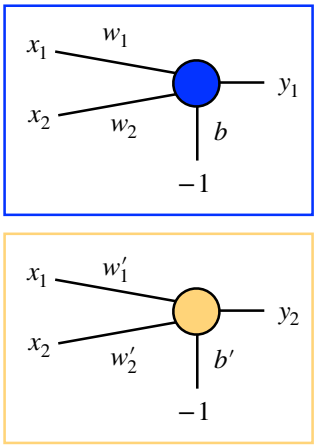


$y = \text{sign}\left(w_2 x_2 + w_1 x_1 - b\right)$

Set of predictors: $\mathrm{H} = \{h_{\mathbf{w},w_0}(\mathbf{x})\} = \text{sign}\left(\left(\sum_{i=1}^{k} w_i x_i\right) - b\right)$



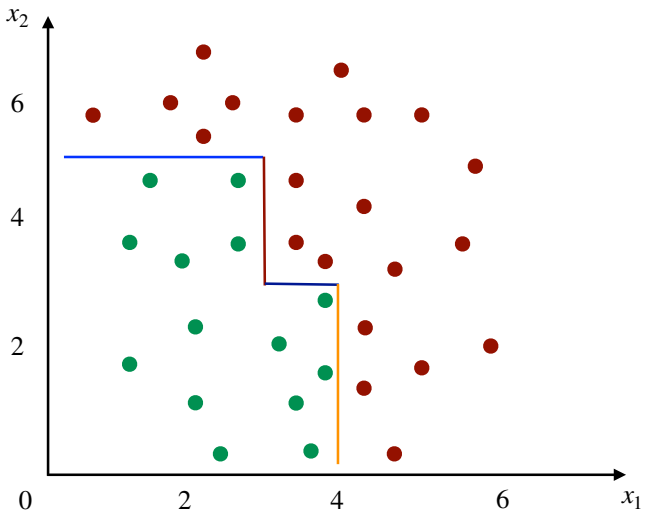
Perceptrons



Neural networks

Combine perceptrons!

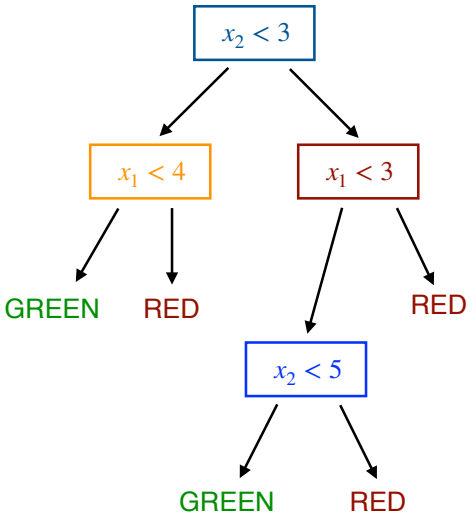
Multi-layer perceptrons



$x_2 < 5$ $x_1 < 3$

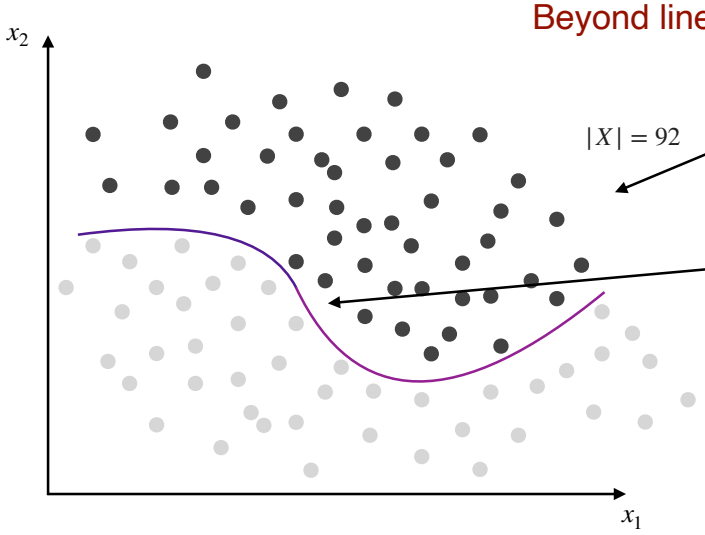
$x_2 < 3$ $x_1 < 4$

Decision trees



Decision trees combine parallel-to-the-axes lines!

Beyond linear predictors



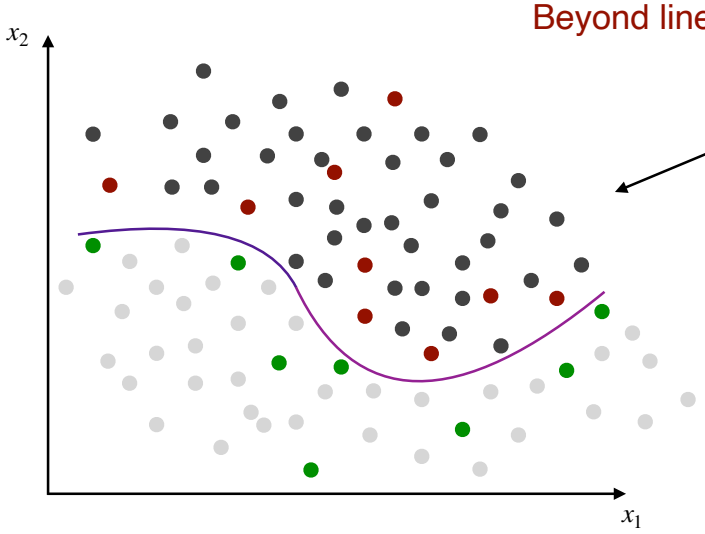
● ● : instances

X : set of instances (domain set)

{GREEN, RED} : set of labels Y

$f : X \to Y$: ideal classifier

Beyond linear predictors



● ● : instances

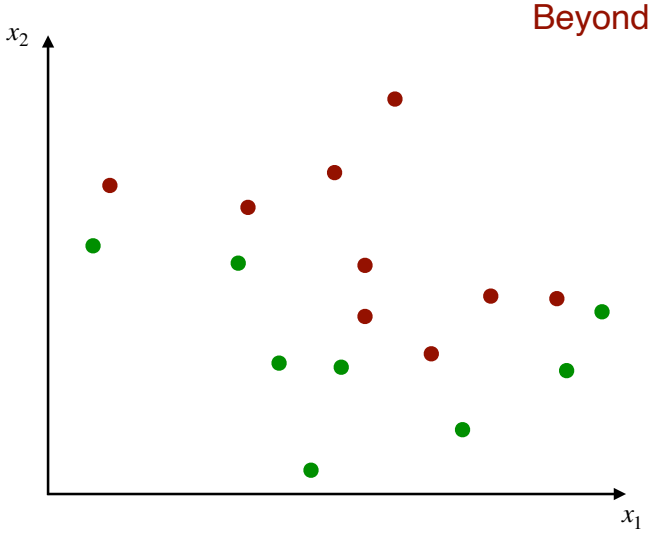
X : set of instances (domain set)

{GREEN, RED} : set of labels Y

$f : X \to Y$: ideal classifier

$\mathbf{x} \sim \mathcal{D}(X)$

Beyond linear predictors



● ● : instances

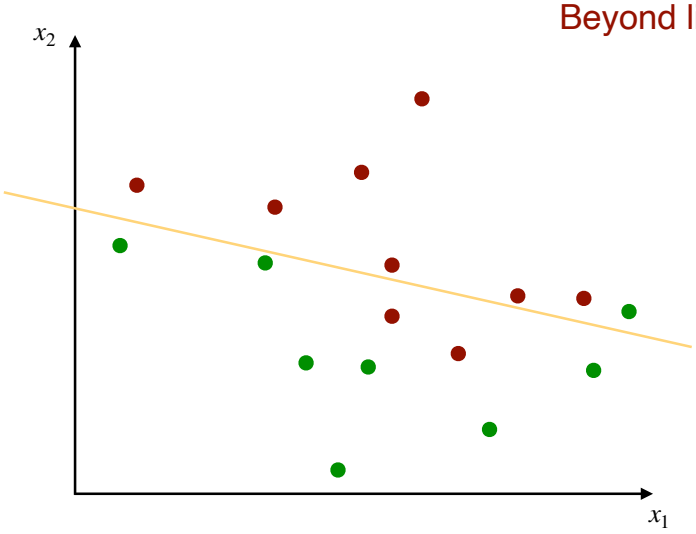
X : set of instances (domain set)

{GREEN, RED} : set of labels Y

$\mathbf{x} \sim \mathcal{D}(X)$

Dataset $\mathbb{D}$

Beyond linear predictors



● ● : instances

X : set of instances (domain set)

{GREEN, RED} : set of labels Y

$\mathbf{x} \sim \mathcal{D}(X)$

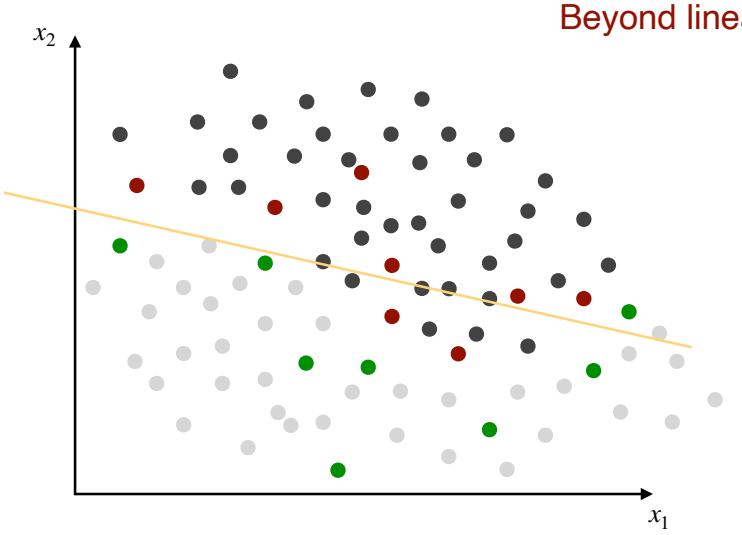
Dataset $\mathbb{D}$ (17 items)

$$L_{\mathbb{D}}(h) = \frac{\left| i \in \mathbb{N}_n : h(\mathbf{x}_i) \neq y_i \right|}{|n|} = \frac{3}{17}$$

the empirical error

Set of predictors: $\mathrm{H} = \{h_{\mathbf{w},w_0}(\mathbf{x})\} = \mathrm{sign}\left(\left(\sum_{i=1}^{k} w_i x_i\right) - b\right)$

Beyond linear predictors



● ● : instances

X : set of instances (domain set)

{GREEN, RED} : set of labels Y

● ●

$f : X \to Y$: ideal classifier

$\mathbf{x} \sim \mathcal{D}(X)$

Dataset $\mathbb{D}$ (17 items)

$$L_{\mathbb{D}}(h) = \frac{\left| i \in \mathbb{N}_n : h(\mathbf{x}_i) \neq y_i \right|}{|n|} = \frac{3}{17}$$

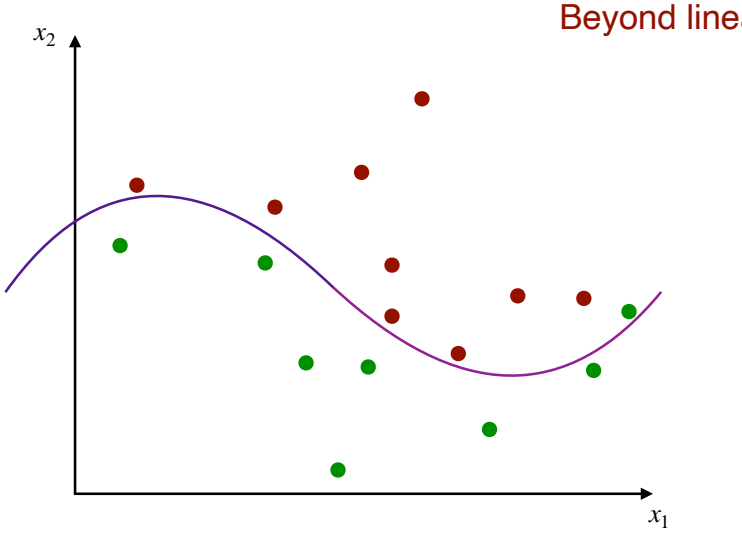
the empirical error

Set of predictors: $\mathrm{H} = \{h_{\mathbf{w},w_0}(\mathbf{x})\} = \mathrm{sign}\left(\left(\sum_{i=1}^{k} w_i x_i\right) - b\right)$

$$L_f(h) = \frac{\left| x \in X : h(x) \neq f(x) \right|}{|X|} = \frac{8}{92}$$

the true error

Beyond linear predictors



● ● : instances

X : set of instances (domain set)

{GREEN, RED} : set of labels Y

● ●

$f : X \to Y$: ideal classifier

$\mathbf{x} \sim \mathcal{D}(X)$

Dataset $\mathbb{D}$ (17 items)

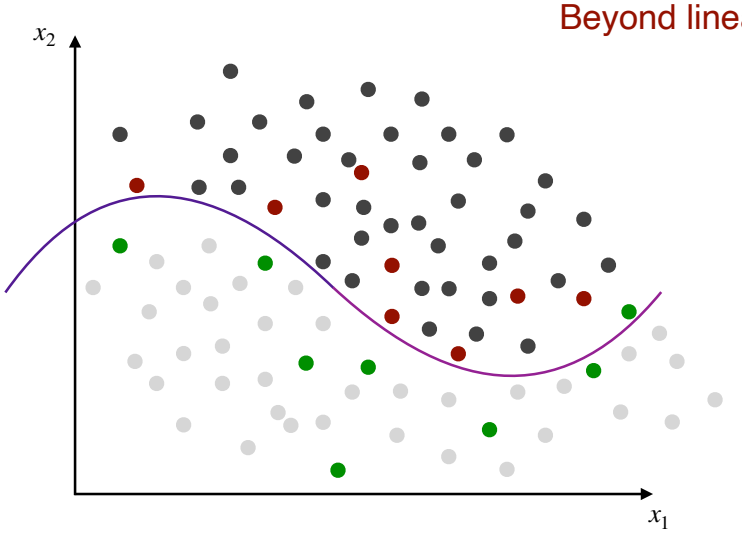
$$L_{\mathbb{D}}(h) = \frac{\left| i \in \mathbb{N}_n : h(\mathbf{x}_i) \neq y_i \right|}{|n|} = \frac{1}{17}$$

the empirical error

Set of predictors: $\mathrm{H} = \{h_{\mathbf{w},w_0}(\mathbf{x})\} = \mathrm{sign}\left(\left(\sum_{m,n\in\mathbb{N}} w_{mn} x_1^m x_2^n\right) - b\right)$

m, n small

Beyond linear predictors



● ● : instances

X : set of instances (domain set)

{GREEN, RED} : set of labels Y

$f : X \to Y$: ideal classifier

$\mathbf{x} \sim \mathcal{D}(X)$

Dataset $\mathbb{D}$ (17 items)

$$L_{\mathbb{D}}(h) = \frac{\left| i \in \mathbb{N}_n : h(\mathbf{x}_i) \neq y_i \right|}{|n|} = \frac{1}{17}$$

the empirical error

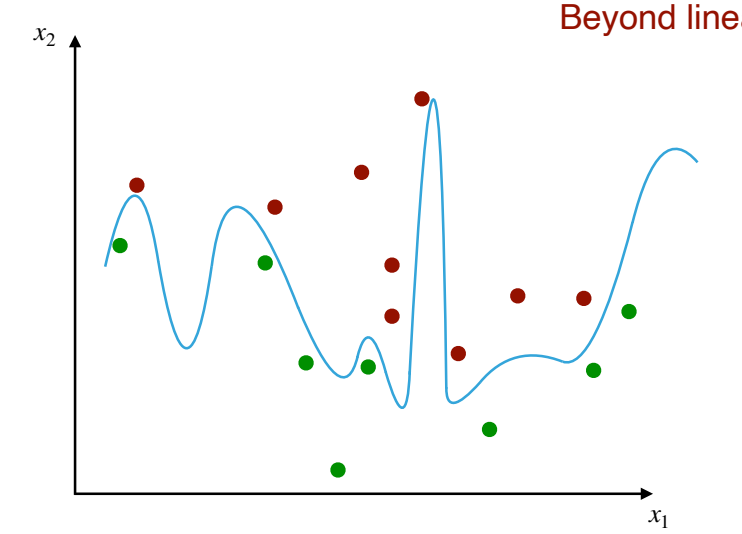
Set of predictors: $\mathrm{H} = \{h_{\mathbf{w},w_0}(\mathbf{x})\} = \text{sign}\left(\left(\sum_{m,n\in\mathbb{N}} w_{mn} x_1^m x_2^n\right) - b\right)$

m, n small

$$L_f(h) = \frac{\left| x \in X : h(x) \neq f(x) \right|}{|X|} = \frac{1}{92}$$

the true error

Beyond linear predictors



● ● : instances

X : set of instances (domain set)

{GREEN, RED} : set of labels Y

$f : X \to Y$: ideal classifier

$\mathbf{x} \sim \mathcal{D}(X)$

Dataset $\mathbb{D}$ (17 items)

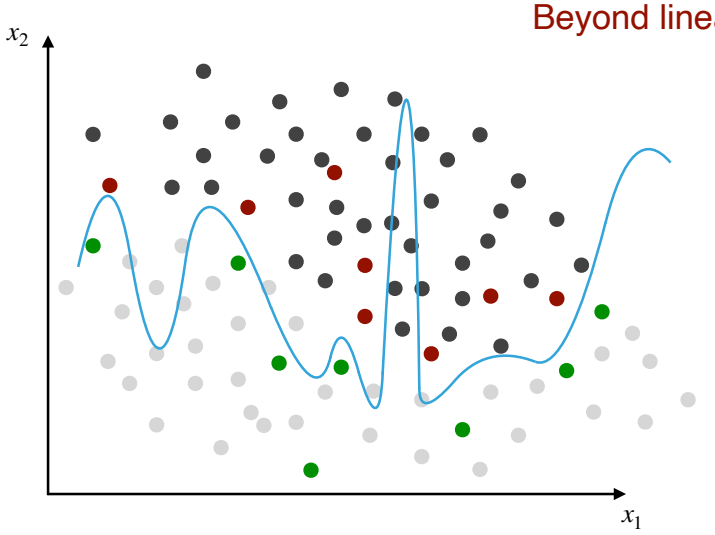
$$L_{\mathbb{D}}(h) = \frac{\left| i \in \mathbb{N}_n : h(\mathbf{x}_i) \neq y_i \right|}{|n|} = \frac{0}{17}$$

the empirical error

Set of predictors: $\mathrm{H} = \{h_{\mathbf{w},w_0}(\mathbf{x})\} = \text{sign}\left(\left(\sum_{m,n\in\mathbb{N}} w_{mn} x_1^m x_2^n\right) - b\right)$

m, n large

Beyond linear predictors



● ● : instances

X : set of instances (domain set)

{GREEN, RED} : set of labels Y

$f : X \to Y$: ideal classifier

$\mathbf{x} \sim \mathscr{D}(X)$

Dataset $\mathbb{D}$ (17 items)

$$L_{\mathbb{D}}(h) = \frac{\left| i \in \mathbb{N}_n : h(\mathbf{x}_i) \neq y_i \right|}{|n|} = \frac{0}{17}$$

the empirical error

Set of predictors: $\mathrm{H} = \{h_{\mathbf{w},w_0}(\mathbf{x})\} = \text{sign}\left(\left(\sum_{m,n\in\mathbb{N}} w_{mn} x_1^m x_2^n\right) - b\right)$

m, n large

$$L_f(h) = \frac{\left| x \in X : h(x) \neq f(x) \right|}{|X|} = \frac{10}{92}$$

the true error

Overfitting!

Why traditional machine learning rocks with tabular data, in practice?

ML engineers can deal with overfitting because:

Features are handcrafted, so they are a result of a priori knowledge about the problem

Datasets usually come from real life, so they capture the complexities of reality $\left(\text{e.g. } \mathbf{x} \sim \mathscr{D}(X)\right)$

Because of the two above facts, datasets can nicely feed data analysis tools

⇒ ML engineers can use data engineering and feature engineering to understand the problem

⇒ ML engineers can train themselves before training neural networks!

⇒ Therefore, ML engineers can understand how complex the ML model should be and what are the complexities of the problem that the model should deal with

Train the right model
with the right data is
not a very difficult task!

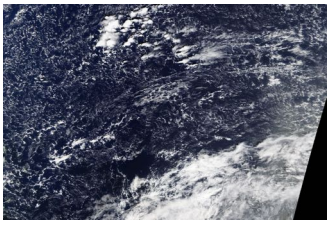
features

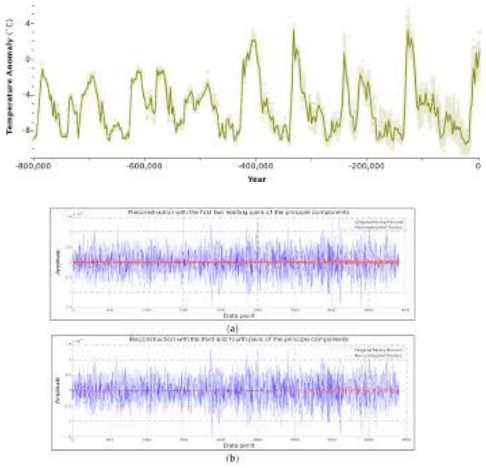
labels

items

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	90	FALSE	No
Sunny	Hot	87	TRUE	No
Overcast	Hot	93	FALSE	Yes
Rainy	Mild	89	FALSE	Yes
Rainy	Cool	79	FALSE	Yes
Rainy	Cool	59	TRUE	No
Overcast	Cool	77	TRUE	Yes
Sunny	Mild	91	FALSE	No
Sunny	Cool	68	TRUE	Yes
Rainy	Mild	80	FALSE	Yes
Sunny	Mild	72	TRUE	Yes
Overcast	Mild	96	TRUE	Yes
Overcast	Hot	74	FALSE	Yes
Rainy	Mild	97	TRUE	No

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	90	FALSE	No
Sunny	Hot	87	TRUE	No
Overcast	Hot	93	FALSE	Yes
Rainy	Mild	89	FALSE	Yes
Rainy	Cool	79	FALSE	Yes
Rainy	Cool	59	TRUE	No
Overcast	Cool	77	TRUE	Yes
Sunny	Mild	91	FALSE	No
Sunny	Cool	68	TRUE	Yes
Rainy	Mild	80	FALSE	Yes
Sunny	Mild	72	TRUE	Yes
Overcast	Mild	96	TRUE	Yes
Overcast	Hot	74	FALSE	Yes
Rainy	Mild	97	TRUE	No





000
AWUS81 KCTP 071512
RWSCTP
PAZ004>006-010>012-017>019-024>028-033>037-041-042-045-046-049>053-056>059-063>066-080515-

Regional Weather Summary
National Weather Service State College PA
1112 AM EDT Sun May 7 2023

Rain and perhaps a few thunderstorms will move in from the west this afternoon, cross the entire state during the day into the evening hours. The heavier rain will be south of I-80, but even the northern tier will get wet. Expect the steadier rain to move off to the southeast by evening hours, but scattered showers will hold on into the night, mainly over the southwestern and northeastern corners of the Commonwealth.

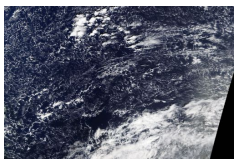
Most of the day on Monday looks dry for the balance of the state. Showers will develop over the west later in the day. Expect a large patch of showers and thunderstorms to cross the state Monday night.

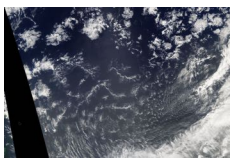
Highs today will be very near normals with values ranging from the mid 60s in the Alleghenies to mid 70s in the southeast. Nighttime lows will only dip to 50 degrees along the northern border and Lake Erie, and stay closer to 60 degrees along the southern border. Highs on Monday will be in the mid 60s in Erie and mid 70s in Philadelphia.

$$

Machine learning for raw data

Max Planck Institute of Meteorology





~10.000 images

Build a model to classify cloud organization patterns from satellite images

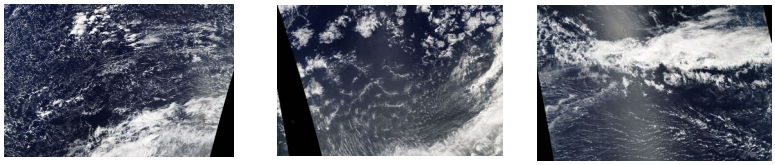
There are many ways in which clouds can organize, but the boundaries between different forms of organization are murky.

This makes it challenging to build traditional rule-based algorithms to separate cloud features.

The human eye, however, is really good at detecting features—such as clouds that resemble flowers.

Traditional machine learning for raw data?

Max Planck Institute of Meteorology

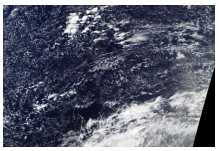


FISH FLOWER GRAVEL SUGAR

Features (MPEG7 descriptors)

Dominant Color
ScalableColor
ColorLayout
ColorStructure
ColorTemperature
GofGopColor
HomogeneousTexture
TextureBrowsing
EdgeHistogram
RegionShape
ContourShape
Perceptual3DShape
…

The Motion Pictures Expert Group



$[x_1, x_2, \ldots, x_n]$

Raw data (images) → Tabular data (features)

Image features (~10^2)

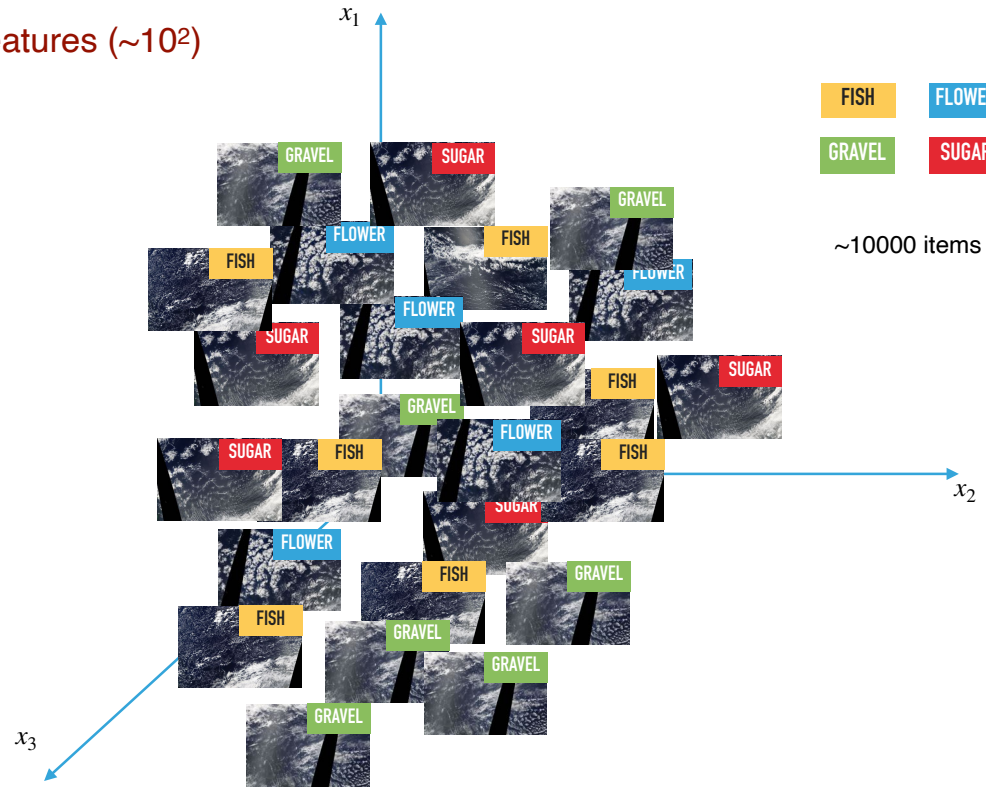


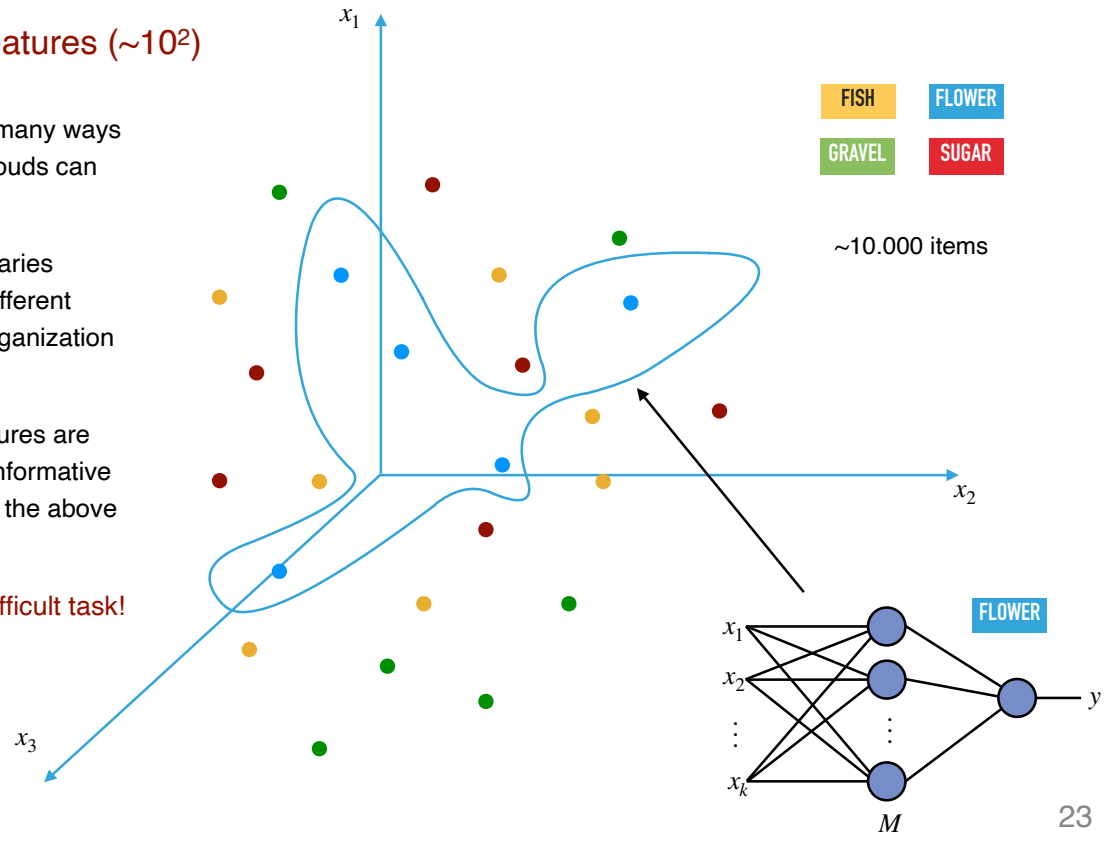
Image features (~10^2)

There are many ways in which clouds can organize

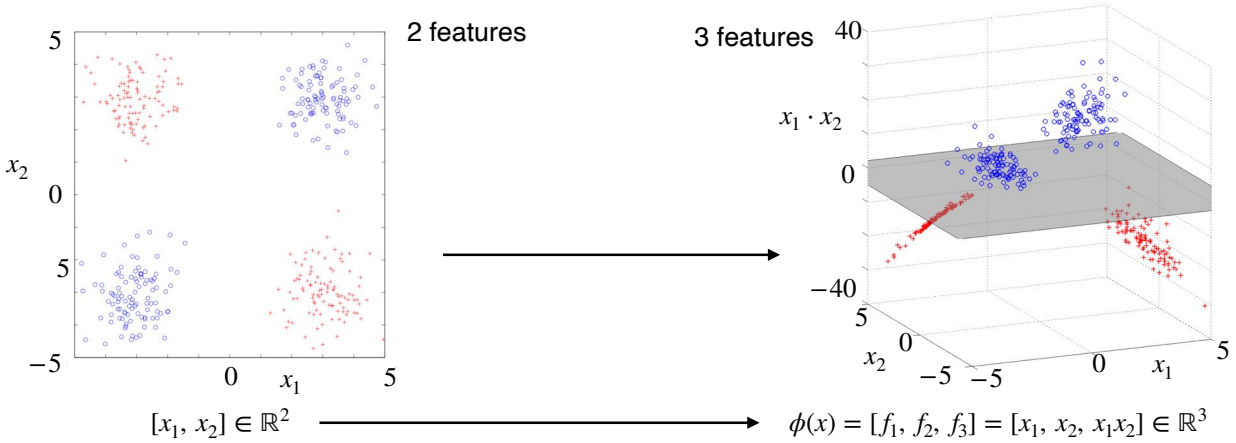
The boundaries between different forms of organization are murky

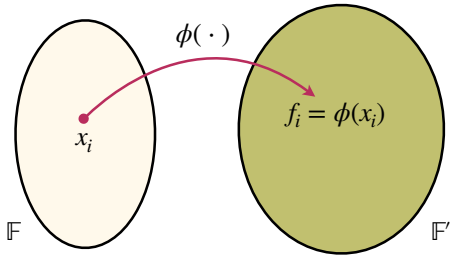
Image features are not really informative for none of the above

This is a difficult task!



Automatically find a convenient feature space





Changing the feature space can make the problem easier

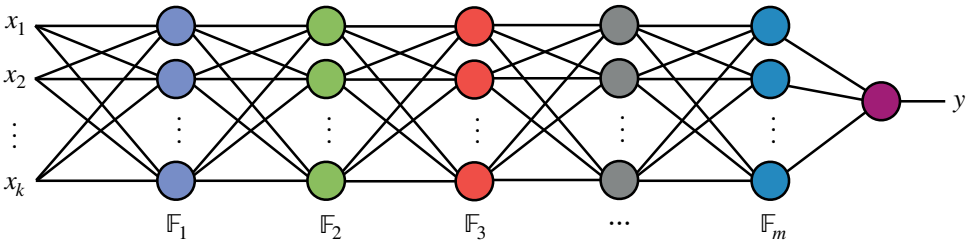
Automatically find a convenient feature space

$$X \longrightarrow Y$$

$$X \longrightarrow \mathbb{F}_1 \longrightarrow \mathbb{F}_2 \longrightarrow \mathbb{F}_3 \longrightarrow \cdots \longrightarrow \mathbb{F}_m \longrightarrow Y$$

Change the feature space arbitrary many times until the mapping is easy

This is the idea behind deep learning



However

The model has many more parameters $\Rightarrow$

We need more data to train the model

We need computational power to train the model

The model is opaque, it works like a black-box

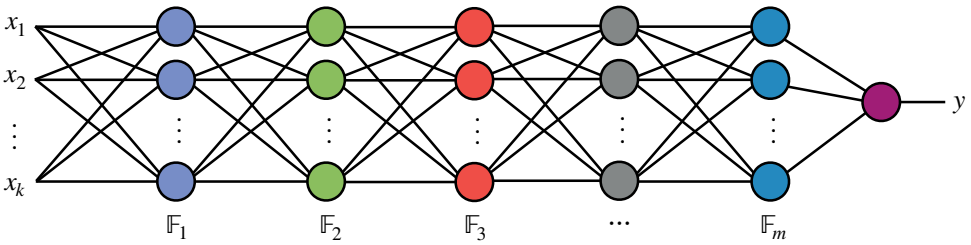
Automatically find a convenient feature space

$$X \longrightarrow Y$$

$$X \longrightarrow \mathbb{F}_1 \longrightarrow \mathbb{F}_2 \longrightarrow \mathbb{F}_3 \longrightarrow \cdots \longrightarrow \mathbb{F}_m \longrightarrow Y$$

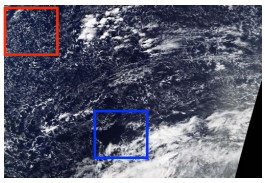
Change the feature space arbitrary many times until the mapping is easy

This is the idea behind deep learning

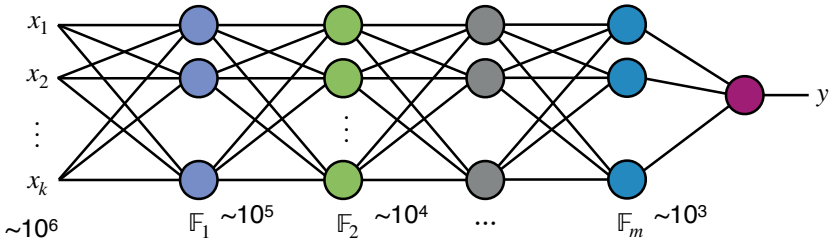


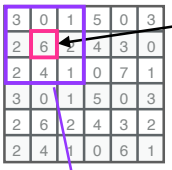
And what is the input?

Convolutional neural networks (CNNs)



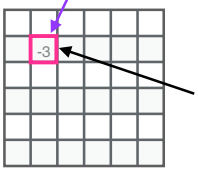
Pixels



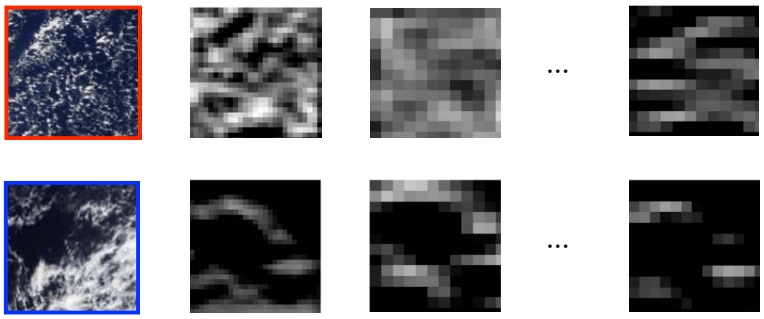


Source pixel

Convolutional filter



Destination pixel



CNNs learn increasingly sophisticated features as they change the layers, finding patterns in the input images

Representation learning

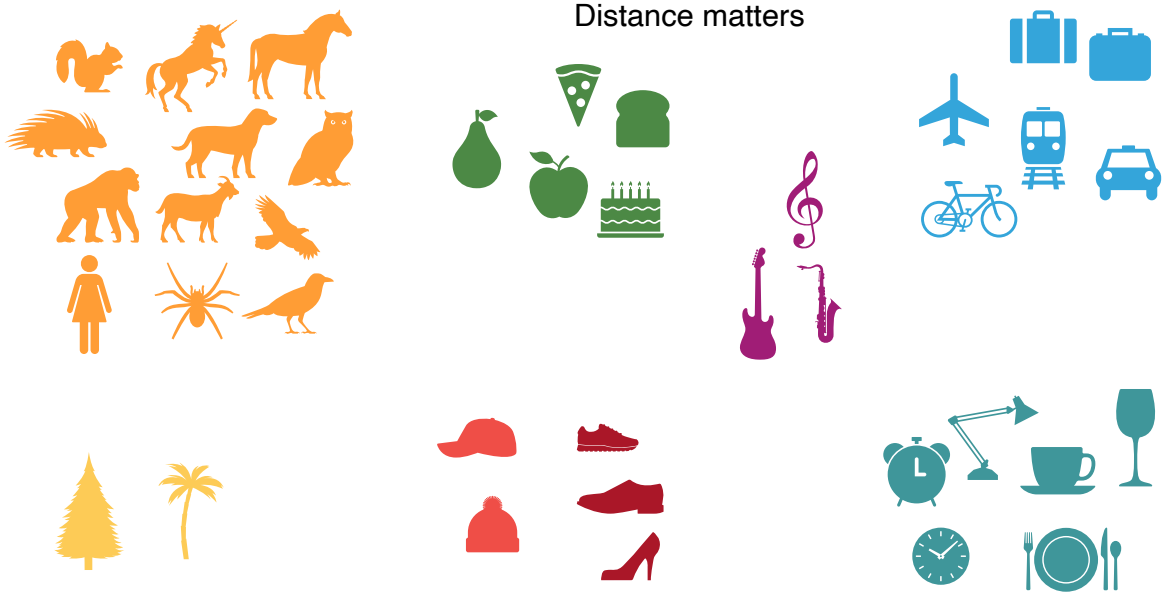
Can we use sophisticated feature spaces to represent and distinguish different things?

Representation learning

Can we use sophisticated feature spaces to represent and distinguish different things?

Representation learning

Can we use sophisticated feature spaces to represent and distinguish different things?



Representation learning

Distance in the feature space should be smaller for semantically similar things!

David Hilbert
(1862-1943)

$\phi(\cdot) = [x_1, x_2, \ldots, x_k]$, k arbitrary large

$\phi(\text{horse}) = [x_1, x_2, \ldots, x_k]$

$\phi(\text{dog}) = [x'_1, x'_2, \ldots, x'_k]$

$\text{dist}\left(\phi(\text{horse}), \phi(\text{dog})\right) = d_1$

$d_1 < d_2$

$\phi(\text{horse}) = [x_1, x_2, \ldots, x_k]$

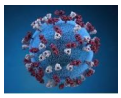
$\phi(\text{guitar}) = [x''_1, x''_2, \ldots, x''_k]$

$\text{dist}\left(\phi(\text{horse}), \phi(\text{guitar})\right) = d_2$

Representation learning

Is it possible to find a feature space that can semantically distinguish all things?




DNA carries genetic information for the development and functioning of an organism

⇒ DNA distinguishes all organisms

Therefore, it is possible!

Representation learning

What problems can we solve using a sophisticated feature representation?

Classification

Problem

Given a dataset $\mathbb{D} = \{\langle x_i, y_i \rangle\}_{i=1}^{n}$

Given an $x \in X$ (an instance of the domain set)

Find the label y of x

Solving Approach

Find the feature represenations $[f_1^{(i)}, f_2^{(i)}, \ldots, f_k^{(i)}], i = 1,2,...,n$ of all the items $x_1, x_2, \ldots, x_n$

Find the feature represenation $[f_1, f_2, \ldots, f_k]$ of x

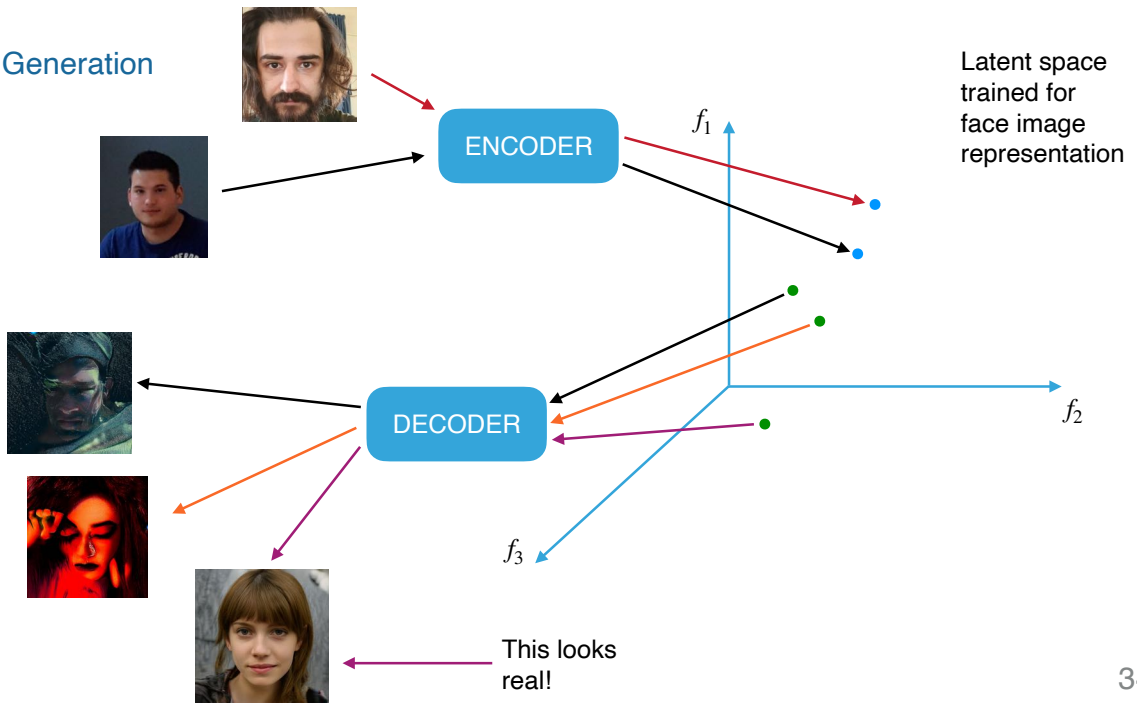
Find the minimal distance between $[f_1, f_2, \ldots, f_k]$ and $[f_1^{(i)}, f_2^{(i)}, \ldots, f_k^{(i)}], i = 1,2,...,n$

Predict the label x as the label of the item x_i with the minimal distance of feature vectors

Representation learning

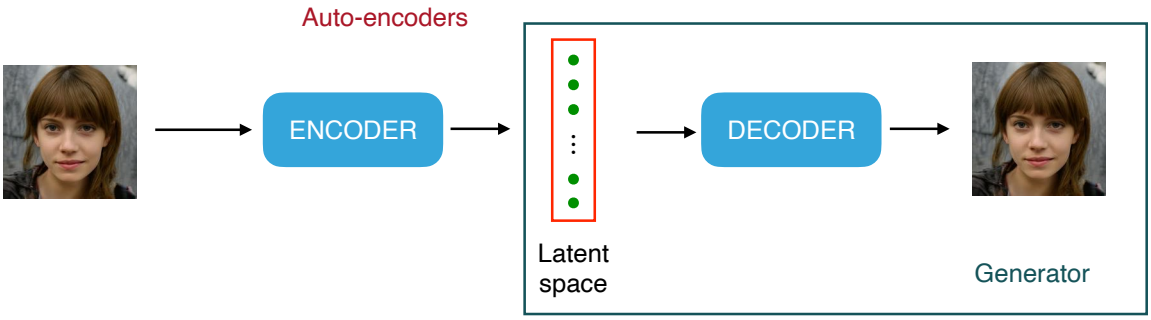
What problems can we solve using a sophisticated feature representation?

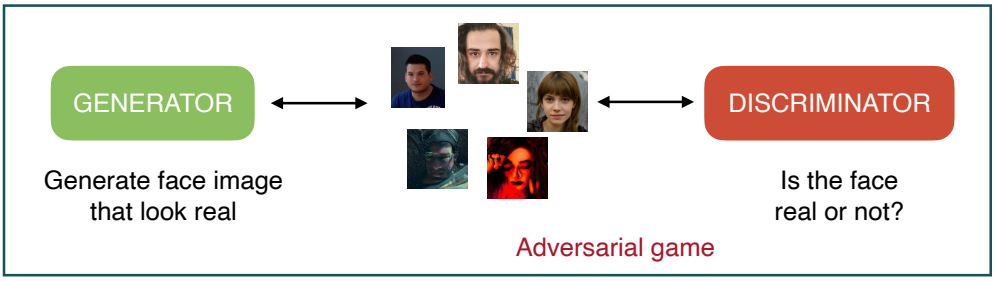
Generation



Representation learning

Can we generate images of “look-real” faces?





Word embeddings

000
AWUS81 KCTP 071512
RWSCTP
PAZ004>006-010>012-017>019-024>028-033>037-041-042-045-046-
049>053-056>059-063>066-080515-

Regional Weather Summary
National Weather Service State College PA
1112 AM EDT Sun May 7 2023

Rain and perhaps a few thunderstorms will move in from the west this afternoon, cross the entire state during the day into the evening hours. The heavier rain will be south of I-80, but even the northern tier will get wet. Expect the steadier rain to move off to the southeast by evening hours, but scattered showers will hold on into the night, mainly over the southwestern and northeastern corners of the Commonwealth.

Most of the day on Monday looks dry for the balance of the state. Showers will develop over the west later in the day. Expect a large patch of showers and thunderstorms to cross the state Monday night.

Highs today will be very near normals with values ranging from the mid 60s in the Alleghenies to mid 70s in the southeast. Nighttime lows will only dip to 50 degrees along the northern border and Lake Erie, and stay closer to 60 degrees along the southern border. Highs on Monday will be in the mid 60s in Erie and mid 70s in Philadelphia.

$$

Can we do the same with texts?

Find a feature space for words

How many words?

We use ~50.000 words

google web crawl gives ~13 million words

Compute the feature representation of every word to find the similarity of words and then the similarity of texts

Can we find a good feature space?

Why not one feature for each word? (one-hot representation)

Word embeddings

000
AWUS81 KCTP 071512
RWSCTP
PAZ004>006-010>012-017>019-024>028-033>037-041-042-045-04
6-
049>053-056>059-063>066-080515-

Regional Weather Summary
National Weather Service State College PA
1112 AM EDT Sun May 7 2023

Rain and perhaps a few **thunderstorms** will move in from the west this afternoon, cross the entire state during the day into the evening hours. The heavier rain will be south of I-80, but even the northern tier will get wet. Expect the steadier rain to move off to the southeast by evening hours, but scattered showers will hold on into the night, mainly over the southwestern and northeastern corners of the Commonwealth.

Most of the day on Monday looks dry for the balance of the state. Showers will develop over the west later in the day. Expect a large patch of **showers** and thunderstorms to cross the state Monday night.

Highs today will be very near normals with values ranging from the mid 60s in the Alleghenies to mid 70s in the southeast. Nighttime lows will only dip to 50 degrees along the northern border and Lake Erie, and stay closer to 60 degrees along the southern border. Highs on Monday will be in the mid 60s in Erie and mid 70s in Philadelphia.

$$

Thunderstorms

[0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]

Showers

[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0]

sim(Thunderstorms,Showers)=0

"You shall know a word by the company it keeps" – J.R. Firth, 1957

Word embeddings

000
AWUS81 KCTP 071512
RWSCTP
PAZ004>006-010>012-017>019-024>028-033>037-041-042-045-
046-
049>053-056>059-063>066-080515-

Regional Weather Summary
National Weather Service State College PA
1112 AM EDT Sun May 7 2023

Rain and perhaps a few thunderstorms will move in from the west this afternoon, cross the entire state during the day into the evening hours. The heavier rain will be south of I-80, but even the northern tier will get wet. Expect the steadier rain to move off to the southeast by evening hours, but scattered showers will hold on into the night, mainly over the southwestern and northeastern corners of the Commonwealth.

Most of the day on Monday looks dry for the balance of the state. Showers will develop over the west later in the day. Expect a large patch of showers and thunderstorms to cross the state Monday night.

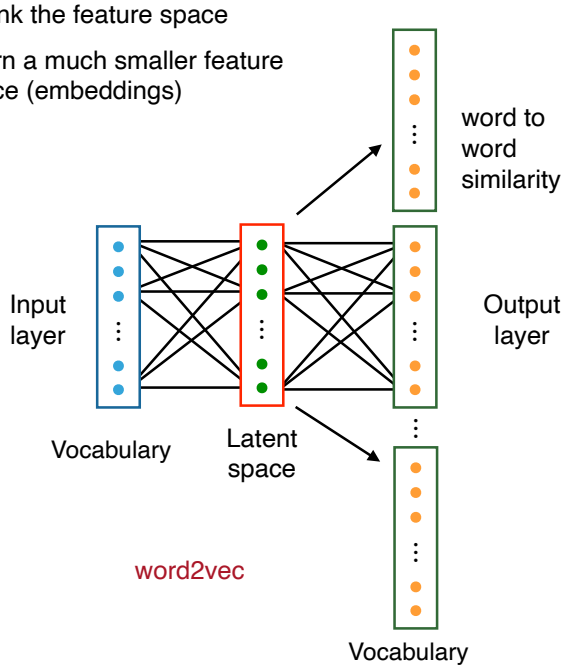
Highs today will be very near normals with values ranging from the mid 60s in the Alleghenies to mid 70s in the southeast. Nighttime lows will only dip to 50 degrees along the northern border and Lake Erie, and stay closer to 60 degrees along the southern border. Highs on Monday will be in the mid 60s in Erie and mid 70s in Philadelphia.

$$

Idea

Shrink the feature space

Learn a much smaller feature space (embeddings)



word2vec