What is Universal AI ?

- The Science of using structured and unstructured **data** to build **models** that lead to **decisions** that provide **value** in all fields.
- Emphasis from **data to value**.
- **Universality** of the audience.
- **Universality** of the material.

Plan

Mornings

- Day I: Predictive AI.
- Day II: Prescriptive AI.
- Day III: Unstructured Data and Deep Learning.

Afternoons

- Exercises + The art of the feasible by subject experts.

Goals

- To **inform** you on the Universality of AI
- To **educate** you in the art of the feasible on AI
- To **inspire** you to deepen your understanding and your education in AI

Vision

- Envision AI as the **universal language** in all fields
- AI to **revolutionize** University Education
- **Horizontal AI** classes that doctors, lawyers, engineers, scientists, … take in the same classroom
- **Vertical education**: AI+physics, AI+medicine, AI+music, Ai+sociology, AI+law, …
- Ultimately to develop Startup Greece with AI as the core Engine

Dimitris Bertsimas

- B.S. Electrical Engineering, NTUA, 1985
- Ph.D. in OR and Applied Math, MIT, 1988
- Professor at MIT Sloan since 1988
- Co-director of the Operations Research Center (2006-2019), Founding Director of MBAn, Current Associate Dean of Business Analytics, MIT
- Member of the US National Academy of Engineering
- Research: AI and Machine Learning, Optimization, Medicine and Healthcare
- Advising: 91 PhD students graduates, 25 current students
- Entrepreneurship: 10 AI companies (4 sold) and 2 nonprofit Foundations

Giorgos Stamou

- Dipl. in Electrical and Computer Engineering, NTUA, 1994
- Ph.D. in Artificial Intelligence, NTUA, 1998
- Professor at NTUA since 2008
- Director of "Artificial Intelligence and Learning Systems Laboratory", Founding director of MSc "Data Science and Machine Learning", NTUA (2018-2022)
- Research: Knowledge Representation and Reasoning, Machine Learning, Applications of AI
- Advising: >80 Diploma and 20 PhD students
- Scientific coordinator or research director of >60 research grants (European, national or private funding)

Plan Today: Predictive AI

- Quality of Wine and Quality of Healthcare
- Framingham Heart Study
- Supreme Court decisions

LECTURE 1 – PREDICTING QUALITY IN WINE AND HEALTHCARE

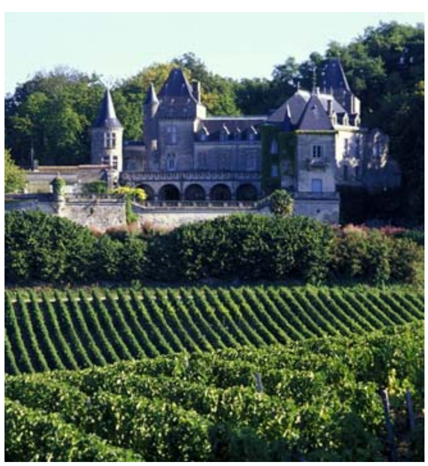
Universal AI

The Statistical Sommelier

An Introduction to Linear Regression

Universal AI

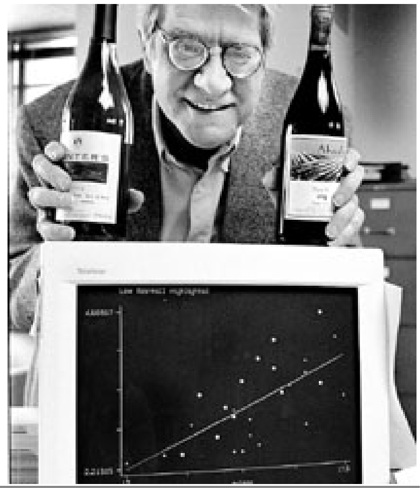
Bordeaux Wine



- Large differences in price and quality between years, although wine is produced in a similar way
- Meant to be aged, so hard to tell if wine will be good when it is on the market
- Expert tasters predict which ones will be good
- Can AI be used to come up with a different system for judging wine?

Predicting the Quality of Wine

- March 1990 - Orley Ashenfelter, a Princeton economics professor, claims he can predict wine quality without tasting the wine



Building a Model

- Ashenfelter used **linear regression**
  - Predicts an outcome variable, or *dependent variable*
  - Predicts using a set of *independent variables*

- Dependent variable: typical price in 1990-1991 wine auctions (approximates quality)

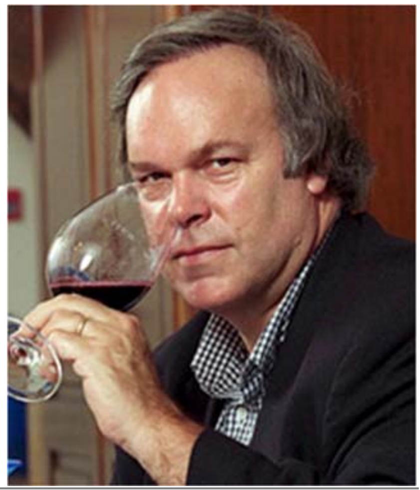
- Independent variables:
  - Age – older wines are more expensive
  - Weather
    - Average Growing Season Temperature
    - Harvest Rain
    - Winter Rain

The Expert's Reaction

Robert Parker, the world's most influential wine expert:

> **"Ashenfelter is an absolute total sham"**

> "rather like a movie critic who never goes to see the movie but tells you how good it is based on the actors and the director"



The Regression Model

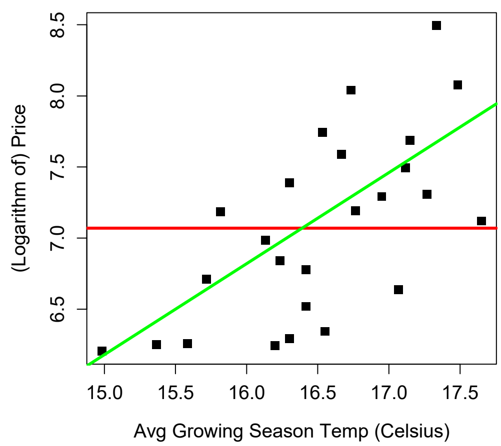
- Multiple linear regression model with k variables

$$y^i = \beta_0 + \beta_1 x_1^i + \beta_2 x_2^i + \ldots + \beta_k x_k^i + \epsilon^i$$

y^i = dependent variable (wine price) for the i^{th} observation
x_j^i = j^{th} independent variable for the i^{th} observation
ϵ^i = error term for the i^{th} observation
β_0 = intercept coefficient
β_j = regression coefficient for the j^{th} independent variable

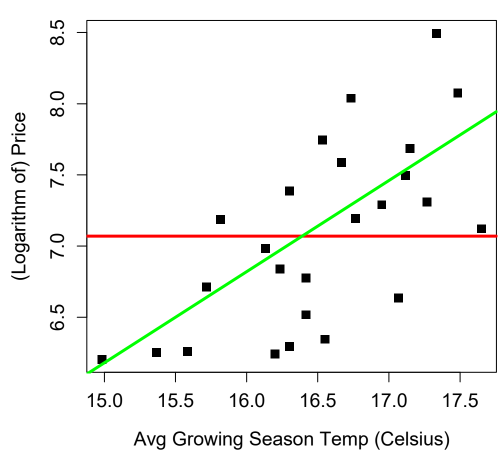
- Best model coefficients selected to minimize SSE

R^2



- Compares the best model to a "baseline" model
- The baseline model does not use any variables
  - Predicts same outcome (price) regardless of the independent variable (temperature)

R^2



SSE = 5.89

SST = 10.15

$$R^2 = 1 - \frac{SSE}{SST}$$

$$R^2 = 1 - \frac{5.89}{10.15}$$

$$R^2 = 0.42$$

Selecting Variables

- Not all available variables should be used
  - Each new variable requires more data
  - Causes *overfitting:* high R^2 on data used to create model, but bad performance on unseen data
- Check for significance
- Check for multicollinearity

The Results

- **Parker:**
  - 1986 is "very good to sometimes exceptional"
- **Ashenfelter:**
  - 1986 is mediocre
  - 1989 will be "the wine of the century" and 1990 will be even better!
- In wine auctions,
  - 1989 sold for more than twice the price of 1986
  - 1990 sold for even higher prices!
- Later, Ashenfelter predicted 2000 and 2003 would be great
- Parker has stated that "2000 is the greatest vintage Bordeaux has ever produced"

The Analytics Edge

- A linear regression model with only a few variables can predict wine prices well
- In many cases, outperforms wine experts' opinions
- A quantitative approach to a traditionally qualitative problem

PREDICTING QUALITY IN HEALTHCARE

An Introduction to Logistic Regression

Univeral AI

Ask the Experts!

- Critical decisions are often made by people with expert knowledge
- Healthcare Quality Assessment
  - Good quality care educates patients and controls costs
  - Need to assess quality for proper medical interventions
  - No single set of guidelines for defining quality of healthcare
  - Health professionals are experts in quality of care assessment

Experts are Human

- Experts are limited by memory and time
- Healthcare Quality Assessment
  - Expert physicians can evaluate quality by examining a patient's records
  - This process is time consuming and inefficient
  - Physicians cannot assess quality for millions of patients

Replicating Expert Assessment

- Can we develop analytical tools that replicate expert assessment on a large scale?
- Learn from expert human judgment
  - Develop a model, interpret results, and adjust the model
- Make predictions/evaluations on a large scale
- Healthcare Quality Assessment
  - Let's identify poor healthcare quality using analytics

Claims Data

Medical Claims

Diagnosis, Procedures, Doctor/Hospital, Cost

Pharmacy Claims

Drug, Quantity, Doctor, Medication Cost

- Electronically available
- Standardized

- Not 100% accurate
- Under-reporting is common
- Claims for hospital visits can be vague

Creating the Dataset – Claims Samples

Claims Sample

- Large health insurance claims database
- Randomly selected 131 diabetes patients
- Ages range from 35 to 55
- Costs $10,000 – $20,000
- September 1, 2003 – August 31, 2005

Creating the Dataset – Expert Review

Claims Sample

Expert Review

- Expert physician reviewed claims and wrote descriptive notes:

"Ongoing use of narcotics"

"Only on Avandia, not a good first choice drug"

"Had regular visits, mammogram, and immunizations"

"Was given home testing supplies"

Creating the Dataset – Expert Assessment

Claims Sample

Expert Review

Expert Assessment

- Rated quality on a two-point scale (poor/good)

"I'd say **care was poor** – poorly treated diabetes"

"No eye care, but overall I'd say **high quality**"

Creating the Dataset – Variable Extraction

Claims Sample

Expert Review

Expert Assessment

Variable Extraction

- Dependent Variable
  - **Quality of care**
- Independent Variables
  - ongoing use of **narcotics**
  - **only on Avandia**, not a good first choice drug
  - Had **regular visits, mammogram, and immunizations**
  - Was given **home testing supplies**

Creating the Dataset – Variable Extraction

Claims Sample

Expert Review

Expert Assessment

Variable Extraction

- Dependent Variable
  - **Quality of care**
- Independent Variables
  - Diabetes treatment
  - Patient demographics
  - Healthcare utilization
  - Providers
  - Claims
  - Prescriptions

Predicting Quality of Care

- The dependent variable is modeled as a binary variable
  - 1 if low-quality care, 0 if high-quality care
- This is a *categorical variable*
  - A small number of possible outcomes
- Linear regression would predict a continuous outcome
- How can we extend the idea of linear regression to situations where the outcome variable is categorical?
  - Only want to predict 1 or 0
  - Could round outcome to 0 or 1
  - But we can do better with **logistic regression**

Logistic Regression

- Predicts the probability of poor care
  - Denote dependent variable "PoorCare" by y
  - $P(y = 1)$
- Then $P(y = 0) = 1 - P(y = 1)$
- Independent variables $x_1, x_2, \ldots, x_k$
- Uses the Logistic Response Function

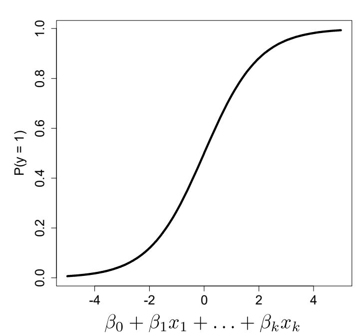
$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}}$$

  - Nonlinear transformation of linear regression equation to produce number between 0 and 1

Understanding the Logistic Function

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}}$$

- Positive values are predictive of class 1
- Negative values are predictive of class 0



Understanding the Logistic Function

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}}$$

- The coefficients are selected to
  - Predict a high probability for the poor care cases
  - Predict a low probability for the good care cases

Understanding the Logistic Function

$$P(y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}}$$

- We can instead talk about Odds (like in gambling)

$$\text{Odds} = \frac{P(y = 1)}{P(y = 0)}$$

- Odds > 1 if y = 1 is more likely
- Odds < 1 if y = 0 is more likely

The Logit

- It turns out that

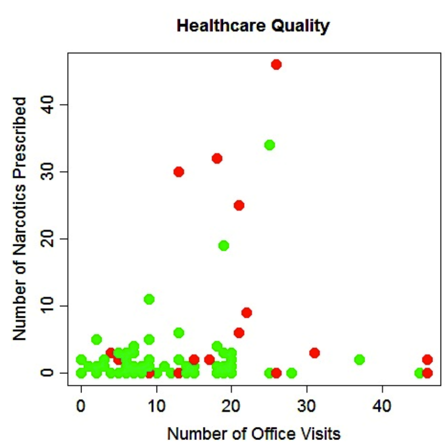
$$\text{Odds} = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_k x_k}$$

$$log(\text{Odds}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

- This is called the "Logit" and looks like linear regression
- Helps us understand the coefficients
  - The bigger the Logit is, the bigger $P(y = 1)$

Model for Healthcare Quality

- Plot of the independent variables
  - Number of Office Visits (OfficeVisits)
  - Number of Narcotics Prescribed (Narcotics)
- Red are poor care
- Green are good care
- Are these variables predictive of good care or poor care?



A Logistic Regression Model

- Used data for 99 patients to build the model (75% of the data)

  Logit = -2.65 + 0.082(OfficeVisits) + 0.076(Narcotics)

- Are higher values in these variables indicative of poor care or good care?
- Now that we have a model, how do we evaluate the quality of the model?

Threshold Value

- The outcome of a logistic regression model is a probability
- Often, we want to make a class prediction to compare with the actual outcome
  - Did this patient receive poor care or good care?
- We can do this using a *threshold value* t
- If P(PoorCare = 1) $\geq$ t, predict poor quality
- If P(PoorCare = 1) < t, predict good quality
- What value should we pick for t?

Threshold Value

- Often selected based on which errors are "better"
- If t is **large**, predict poor care rarely (when P(y=1) is large)
  - More errors where we say good care, but it is actually poor care
  - Detects patients who are receiving the worst care
- If t is **small**, predict good care rarely (when P(y=1) is small)
  - More errors where we say poor care, but it is actually good care
  - Detects all patients who might be receiving poor care
- With no preference between the errors, select t = 0.5
  - Predicts the more likely outcome

Selecting a Threshold Value

Compare actual outcomes to predicted outcomes using a *confusion matrix (classification matrix)*

	Predicted Class = 0	**Predicted Class = 1**
Actual Class = 0	True Negatives (TN) ✓	False Positives (FP) ✗
Actual Class = 1	False Negatives (FN) ✗	True Positives (TP) ✓

- A different threshold value changes the types of errors
- Quantify this trade-off using **sensitivity** and **specificity**
- **Sensitivity = TP/(TP+FN)**
- **Specificity = TN/(TN + FP)**

Selecting a Threshold Value

Which threshold should we select?

Threshold = 0.5

Sensitivity = 10/(10+15) = 0.4

Specificity = 70/(70+4) = 0.95

	Predicted = 0	Predicted = 1
Actual = 0	70	4
Actual = 1	15	10

Threshold = 0.7

Sensitivity = 8/(8+17) = 0.32

Specificity = 73/(73+1) = 0.99

	Predicted = 0	Predicted = 1
Actual = 0	73	1
Actual = 1	17	8

Threshold = 0.2

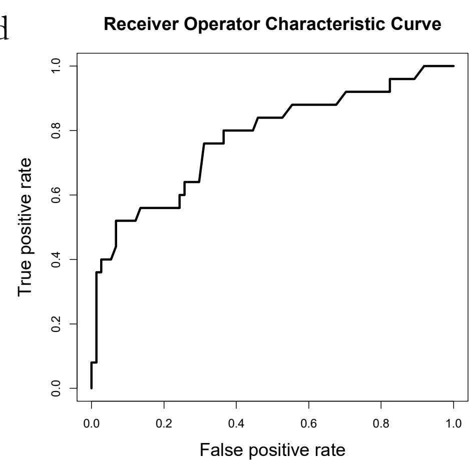
Sensitivity = 16/(16+9) = 0.64

Specificity = 54/(54+20) = 0.73

	Predicted = 0	Predicted = 1
Actual = 0	54	20
Actual = 1	9	16

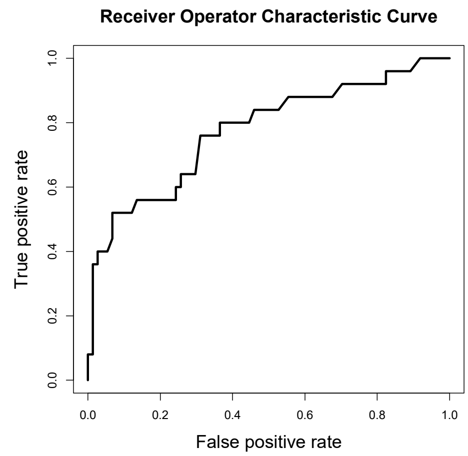
Receiver Operator Characteristic (ROC) Curve

- Helps us pick a threshold
- True positive rate (sensitivity) on y-axis
  - Proportion of poor care caught
- False positive rate
  (1-specificity) on x-axis
  - Proportion of good care labeled as poor care



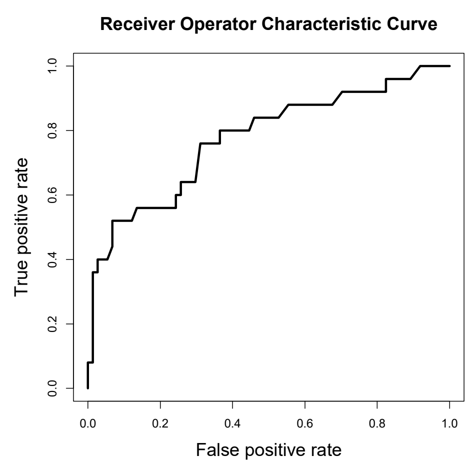
Selecting a Threshold using ROC

- Captures all thresholds simultaneously
- **High threshold**
  - High specificity
  - Low sensitivity
- **Low Threshold**
  - Low specificity
  - High sensitivity



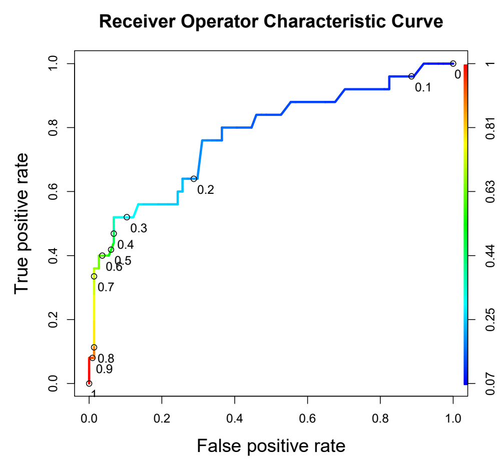
Selecting a Threshold using ROC

- Choose **best threshold** for **best trade off**
  - **cost of failing to detect positives**
  - **costs of raising false alarms**



Selecting a Threshold using ROC

- We'll see in the next lecture how we can generate the ROC curve in R
- We can add colors and labels corresponding to the threshold values

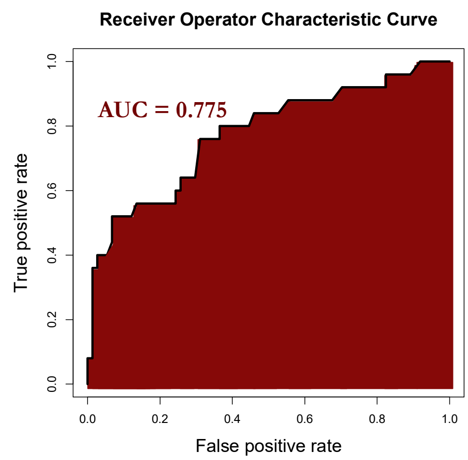


Evaluating the Model

- There are many ways to evaluate the quality of the model
- We'll mainly use two:
  - AUC
  - Accuracy

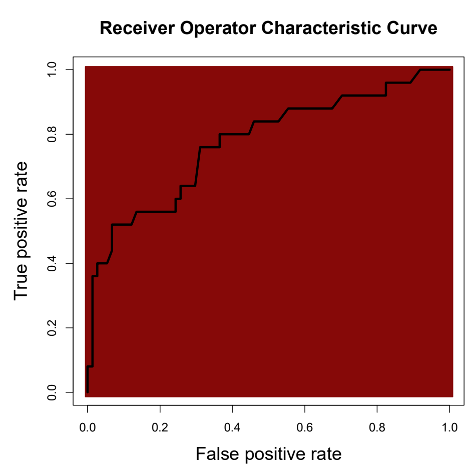
Area Under the ROC Curve (AUC)

- Just take the area under the ROC curve
- Interpretation
  - Given a random positive and negative, proportion of the time you guess which is which correctly



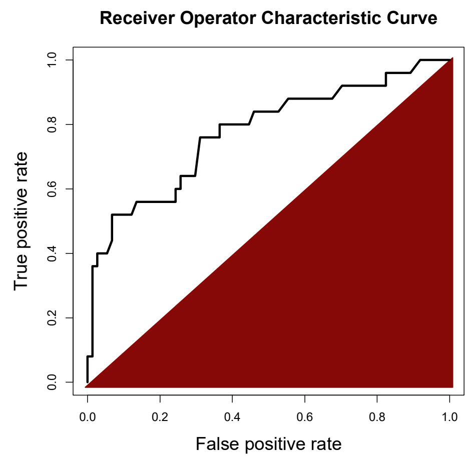
Area Under the ROC Curve (AUC)

- What is a good AUC?
  - Maximum of 1
    (perfect prediction)



Area Under the ROC Curve (AUC)

- What is a good AUC?
  - Maximum of 1
    (perfect prediction)
  - Minimum of 0.5
    (just guessing)



Accuracy

Confusion Matrix:

	Predicted Class = 0	Predicted Class = 1
Actual Class = 0	True Negatives (TN)	False Positives (FP)
Actual Class = 1	False Negatives (FN)	True Positives (TP)

Accuracy = (TN + TP)/(# obs.)

Threshold = 0.5 ➔ Accuracy = 80/99 = 80.8%

Is this good?

A "Baseline Method"

- When we build classification models, we want to compare our model to a simple baseline method
  - Remember that R^2 does this for us in linear regression
- A standard baseline method is to predict the most common outcome
- In this case, 98 patients actually received good care, and 33 patients actually received poor care
- The baseline method would predict good care for everyone, and get an accuracy of 98/131 = 74.8%

Making Predictions

- Just like in linear regression, we want to make predictions on a test set to compute out-of-sample metrics
  - We have 32 patients in our test set
- If we use a threshold value of 0.5, we get the following confusion matrix

	Predicted Good Care	Predicted Poor Care
Actually Good Care	23	1
Actually Poor Care	5	3

- Out-of-sample accuracy of 26/32 = 81.3%

Conclusions

- An expert-trained model can accurately identify diabetics receiving low-quality care
- In practice, the probabilities returned by the logistic regression model can be used to prioritize patients for intervention
- Electronic medical records could be used in the future

The Competitive Edge of Models

- While humans can accurately analyze small amounts of information, models allow larger scalability
- Models do not replace expert judgment
  - Experts can improve and refine the model
- Models can integrate assessments of many experts into one final unbiased and unemotional prediction