Prediction and Linear Regression

Big Data y Machine Learning para Economía Aplicada

Ignacio Sarmiento-Barbieri

Universidad de los Andes

Agenda

Agenda

What is Big Data?

- Classical Statistics (small data?)
  - Extracting the most from a small amount of data
  - Highly structured data, carefully constructed $\rightarrow$ approximates random sampling (expensive, slow) but very good and reliable
- What is Big Data (the 3 V's)?
  - **V**olume (n and k)
  - **V**elocity
  - **V**ariety

What is Machine learning?

Is all about prediction

- Machine learning is a branch of computer science and statistics, tasked with developing algorithms to predict outcomes y from observable variables x.
- The learning part comes from the fact that we don't specify how exactly the computer should predict y from x. This is left as an empirical problem that the computer can "learn".
- In general, this means that we abstract from the underlying model, the approach is pragmatic

What is Machine learning?

Is all about prediction

- Machine learning is a branch of computer science and statistics, tasked with developing algorithms to predict outcomes y from observable variables x.
- The learning part comes from the fact that we don't specify how exactly the computer should predict y from x. This is left as an empirical problem that the computer can "learn".
- In general, this means that we abstract from the underlying model, the approach is pragmatic

"Whatever works, works...."

“Whatever works, works....”

“Whatever works, works....”????

- In many applications, ML techniques can be successfully applied by data scientists with little knowledge of the problem domain.
- For example, the company Kaggle hosts prediction competitions (`www.kaggle.com/competitions`) in which a sponsor provides a data set, and contestants around the world can submit entries, often predicting successfully despite limited context about the problem.

"Whatever works, works...."????

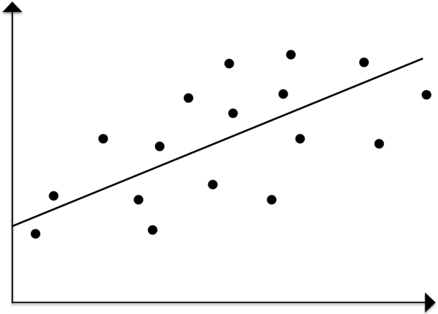
- However, much less attention has been paid to the limitations of pure prediction methods.
- When ML applications are used "off the shelf" without understanding the underlying assumptions or ensuring that conditions like stability are met, then the validity and usefulness of the conclusions can be compromised.
- A deeper question concerns whether a given problem can be solved using only techniques for prediction, or whether statistical approaches to estimating the causal effect of an intervention are required.

Agenda

Prediction vs. Causality: Target

$$y = f(x) + \epsilon \tag{1}$$

$$y = \alpha + \beta x + \epsilon \tag{2}$$



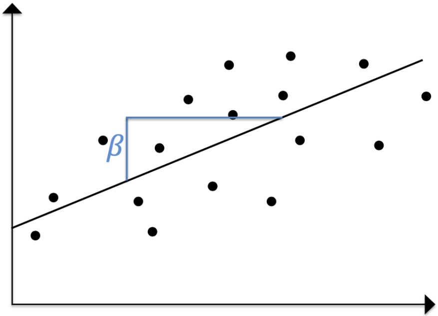
The Causal Paradigm

$$y = f(X) + u \tag{3}$$

- Interest lies on inference
- "Correct" $f()$ to understand how y is affected by X
- Model: Theory, experiment
- Hypothesis testing (std. err., tests)

Prediction vs. Causality: Target

$$y = \alpha + \beta x + \epsilon \tag{4}$$



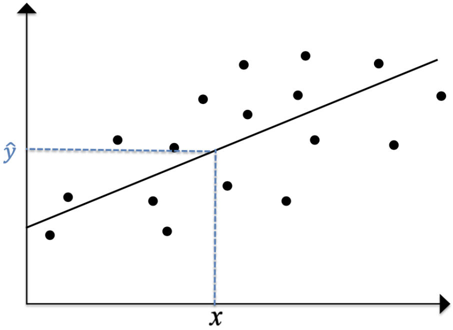
The Predictive Paradigm

$$y = f(X) + u \tag{5}$$

- Interest on predicting y
- "Correct" $f()$ to be able to predict (no inference!)
- Model? We treat $f()$ as a black box, and any approximation $\hat{f}()$ that yields a good prediction is good enough (*Whatever works, works.*).

Prediction vs. Causality: Target

$$y = \underbrace{\alpha + \beta x}_{\hat{y}} + \epsilon \tag{6}$$



Prediction vs. Causality: The garden of the parallel paths?

- We've seen that prediction and causality
  - Answer different questions
  - Serve different purposes
  - Seek different targets
- Different strokes for different folks, or complementary tools in an applied economist's toolkit?

Agenda

Agenda

The basic logic of prediction

Agenda

Prediction error

$$e_j = y_j - \hat{y}_j \tag{7}$$

Minimizing our losses

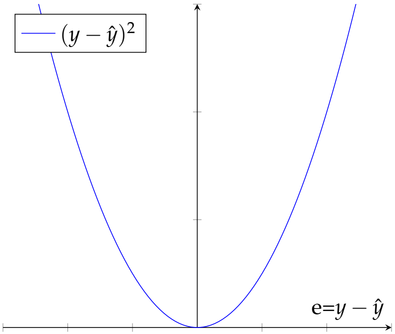
- When making a prediction we want to minimize the prediction errors

$$L(\hat{y}, y) \tag{8}$$

Minimizing our losses

- When making a prediction we want to minimize the prediction errors

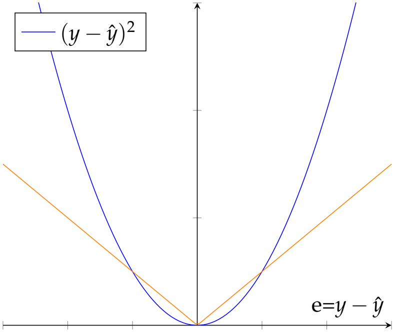
$$L(\hat{y}, y) \tag{8}$$



Minimizing our losses

- When making a prediction we want to minimize the prediction errors

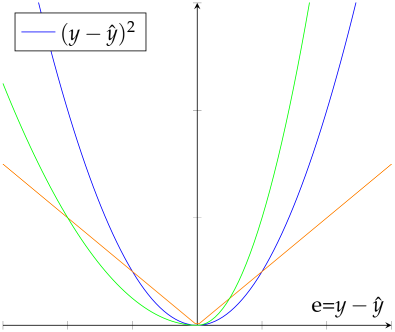
$$L(\hat{y}, y) \tag{9}$$



Minimizing our losses

- When making a prediction we want to minimize the prediction errors

$$L(\hat{y}, y) \tag{10}$$



Minimizing our losses

- We are interested in minimizing our losses
- Can we find the function f^* within a function class $\mathcal{F}$ that has a low expected prediction loss?

Minimizing our losses

- We are interested in minimizing our losses
- Can we find the function f^* within a function class $\mathcal{F}$ that has a low expected prediction loss?
- By conditioning on X, it suffices to minimize the $MSE(f)$ point wise

$$f(x) = argmin_{f^*} E_{Y|X}[(y - f^*)^2 | X = x) \tag{11}$$

Minimizing our losses

- f^* a random variable and we can treat f^* as a constant (predictor)

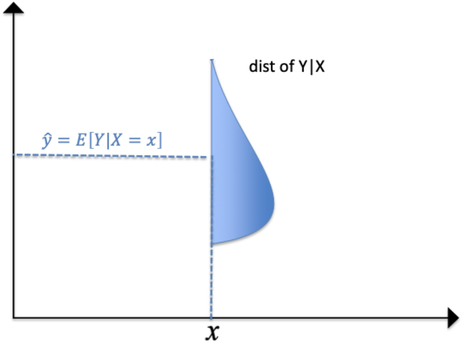
$$min_{f^*} E(y - f^*)^2 = \int (y - f^*)^2 f(y) dy \tag{12}$$

- **Result**: The best prediction of Y at any point $X = x$ is the conditional mean, when best is measured using a square error loss

Minimizing our losses

- **Result**: The best prediction of y at any point $X = x$ is the conditional mean, when best is measured using a square error loss

$$f^* = E[y|X = x] \tag{13}$$



Minimizing our losses

- Prediction problem solved if we knew $f^* = E[y|X = x]$

Minimizing our losses

- Prediction problem solved if we knew $f^* = E[y|X = x]$
- But we have to settle for an estimate: $\hat{f}(x)$

Agenda

Prediction and linear regression

- As economists we know that we can approximate $E[y|X=x]$ with a linear regression

$$f(X) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{14}$$

- The problem boils down to choosing the right functional form
- and finding the coefficients β

Agenda

Choosing the right functional form

photo from https://www.dailydot.com/parsec/batman-1966-labels-tumblr-twitter-vine/

Agenda

Obtaining the coefficients

photo from https://www.dailydot.com/parsec/batman-1966-labels-tumblr-twitter-vine/

Agenda

Review

- This Module: The predictive paradigm and linear regression
  - Machine Learning is all about prediction
  - ML targets something different than causal inference, they can complement each other
  - Linear Regression can approximate $E(y|X)$
  - Inner workings of linear regression
- Next Module: Out of sample prediction. Over-fit, Resampling Methods, Web-scrapping