Week 2 Video 4

Metrics for Regressors

Metrics for Regressors

- Linear Correlation
- MAE/RMSE
- Information Criteria

Linear correlation (Pearson's correlation)

- r(A,B) =
- When A's value changes, does B change in the same direction?
- Assumes a linear relationship

What is a “good correlation”?

- 1.0 – perfect
- 0.0 – none
- -1.0 – perfectly negatively correlated

- In between – depends on the field

What is a “good correlation”?

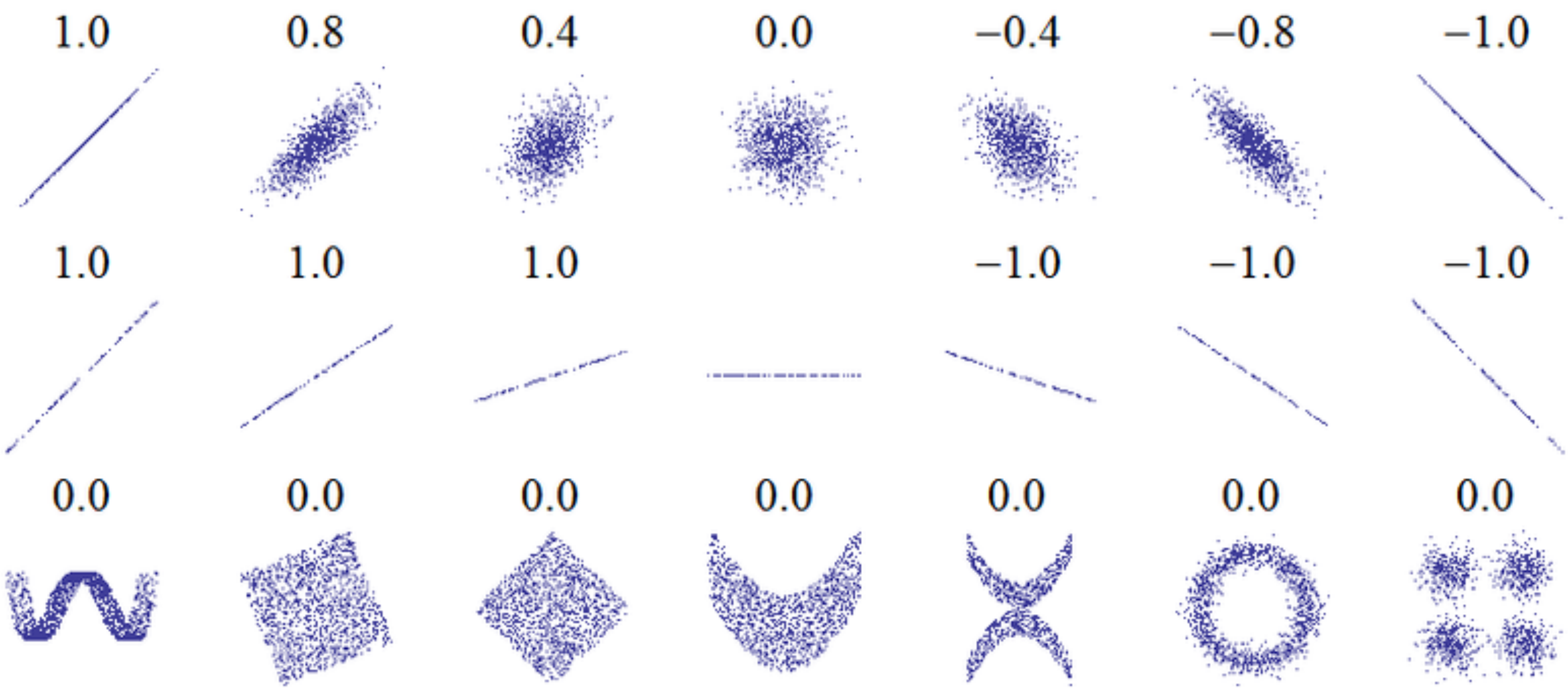
- 1.0 – perfect
- 0.0 – none
- -1.0 – perfectly negatively correlated

- In between – depends on the field
- In physics – correlation of 0.8 is weak!
- In education – correlation of 0.3 is good

Why are small correlations OK in education?

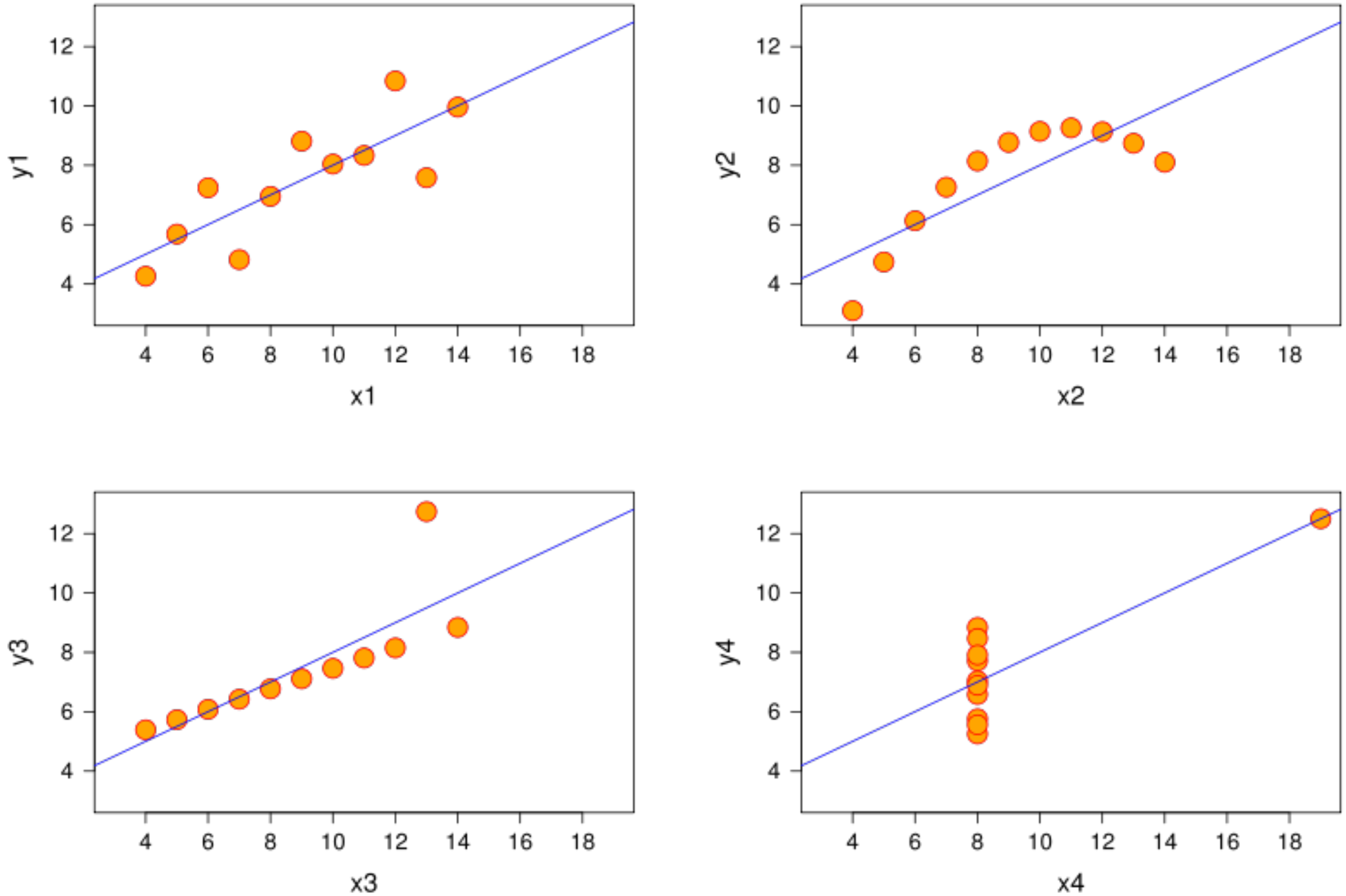
- Lots and lots of factors contribute to just about any dependent measure

Examples of correlation values



From Denis Boigelot, available on Wikipedia

Same correlation, different functions



Anscombe's Quartet

r^2

- The correlation, squared
- Also a measure of what percentage of variance in dependent measure is explained by a model
- If you are predicting A with B,C,D,E
  - r^2 is often used as the measure of model goodness rather than r (depends on the community)

Spearman's Correlation (ρ)

- Rank correlation

- Turn each variable into ranks
- 1 = highest value, 2 = 2nd highest value, 3 = 3rd highest value, and so on

- Then compute Pearson's correlation

- (There's actually an easier formula, but not relevant here)

Spearman's Correlation (ρ)

- Interpreted exactly the same way as Pearson's correlation

- 1.0 – perfect
- 0.0 – none
- -1.0 – perfectly negatively correlated

Why use Spearman's Correlation (ρ)?

- More robust to outliers
- Determines how monotonic a relationship is, not how linear it is

RMSE/MAE

Mean Absolute Error

- Average of

- Absolute value
  (actual value minus predicted value)

Root Mean Squared Error (RMSE)

- Square Root of average of
- (actual value minus predicted value)^2

MAE vs. RMSE

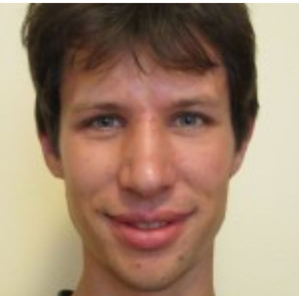
- MAE tells you the average amount to which the predictions deviate from the actual values
  - Very interpretable

- RMSE can be interpreted the same way (mostly) but penalizes large deviation more than small deviation

However

- RMSE is largely preferred to MAE

The example to follow is courtesy of
Radek Pelanek, Masaryk University



Radek's Example

- Take a student who makes correct responses 70% of the time

- And two models
  - Model A predicts 70% correctness
  - Model B predicts 100% correctness

In other words

- 70% of the time the student gets it right
  - Response = 1
- 30% of the time the student gets it wrong
  - Response = 0

- Model A Prediction = 0.7
- Model B Prediction = 1.0

- Which of these seems more reasonable?

MAE

- 70% of the time the student gets it right
  - Response = 1
  - Model A (0.7) Absolute Error = 0.3
  - Model B (1.0) Absolute Error = 0
- 30% of the time the student gets it wrong
  - Response = 0
  - Model A (0.7) Absolute Error = 0.7
  - Model B (1.0) Absolute Error = 1

MAE

- Model A
  - (0.7)(0.3)+(0.3)(0.7)
  - 0.21+0.21
  - 0.42

- Model B
  - (0.7)(0)+(0.3)(1)
  - 0+0.3
  - 0.3

MAE

- Model A
  - (0.7)(0.3)+(0.3)(0.7)
  - 0.21+0.21
  - 0.42

- **Model B is better, according to MAE**
  - (0.7)(0)+(0.3)(1)
  - 0+0.3
  - 0.3

Do you believe it?

- Model A
  - (0.7)(0.3)+(0.3)(0.7)
  - 0.21+0.21
  - 0.42

- **Model B is better, according to MAE**
  - (0.7)(0)+(0.3)(1)
  - 0+0.3
  - 0.3

RMSE

- 70% of the time the student gets it right
  - Response = 1
  - Model A (0.7) Squared Error = 0.09
  - Model B (1.0) Squared Error = 0
- 30% of the time the student gets it wrong
  - Response = 0
  - Model A (0.7) Squared Error = 0.49
  - Model B (1.0) Squared Error = 1

RMSE

- Model A
  - (0.7)(0.09)+(0.3)(0.49)
  - 0.063+0.147
  - 0.21

- Model B
  - (0.7)(0)+(0.3)(1)
  - 0+0.3
  - 0.3

RMSE

- **Model A is better, according to RMSE.**
  - (0.7)(0.09)+(0.3)(0.49)
  - 0.063+0.147
  - 0.21

- Model B
  - (0.7)(0)+(0.3)(1)
  - 0+0.3
  - 0.3

RMSE

- **Model A is better, according to RMSE. Does this seem more reasonable?**
  - (0.7)(0.09)+(0.3)(0.49)
  - 0.063+0.147
  - 0.21

- Model B
  - (0.7)(0)+(0.3)(1)
  - 0+0.3
  - 0.3

Note

- Low RMSE is good
- High Correlation is good

What does it mean?

- Low RMSE/MAE, High Correlation = Good model
- High RMSE/MAE, Low Correlation = Bad model

What does it mean?

- High RMSE/MAE, High Correlation = Model goes in the right direction, but is systematically biased
  - A model that says that adults are taller than children
  - But that adults are 8 feet tall, and children are 6 feet tall

What does it mean?

- Low RMSE/MAE, Low Correlation = Model values are in the right range, but model doesn't capture relative change
  - Particularly common if there's not much variation in data

Information Criteria

BiC

- Bayesian Information Criterion (Raftery, 1995)
- Makes trade-off between goodness of fit and flexibility of fit (number of parameters)
- Formula for linear regression
  - $BiC' = n \log (1 - r^2) + p \log n$
- n is number of students, p is number of variables

BiC'

- Values over 0: worse than expected given number of variables
- Values under 0: better than expected given number of variables

- Can be used to understand significance of difference between models
(Raftery, 1995)

BiC

- Said to be statistically equivalent to k-fold cross-validation for optimal k
- The derivation is… somewhat complex
- BiC is easier to compute than cross-validation, but different formulas must be used for different modeling frameworks
  - No BiC formula available for many modeling frameworks

AIC

- Alternative to BiC

- Stands for
  - An Information Criterion (Akaike, 1971)
  - Akaike's Information Criterion (Akaike, 1974)

- Makes slightly different trade-off between goodness of fit and flexibility of fit (number of parameters)

AIC

- Said to be statistically equivalent to Leave-Out-One-Cross-Validation

AIC or BIC:
Which one should you use?

- <shrug>

All the metrics: Which one should you use?

- "The idea of looking for a single best measure to choose between classifiers is wrongheaded." – Powers (2012)

Next Lecture

- Cross-validation and over-fitting