Week 2 Video 4

Metrics for Regressors

Metrics for Regressors

- Linear Correlation
- MAD/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”?

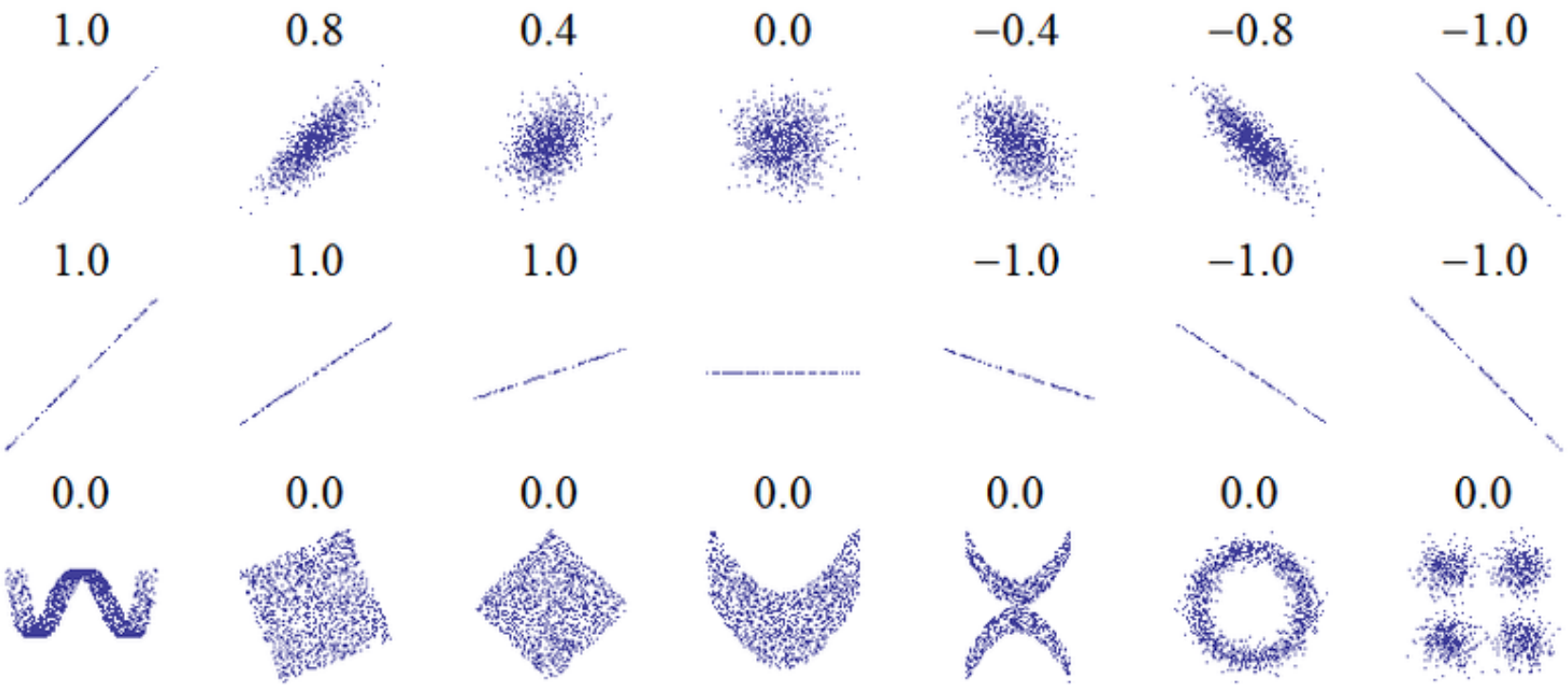
- 1.0 – perfect
- 0.0 – none
- -1.0 – perfectly negatively correlated

- 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?

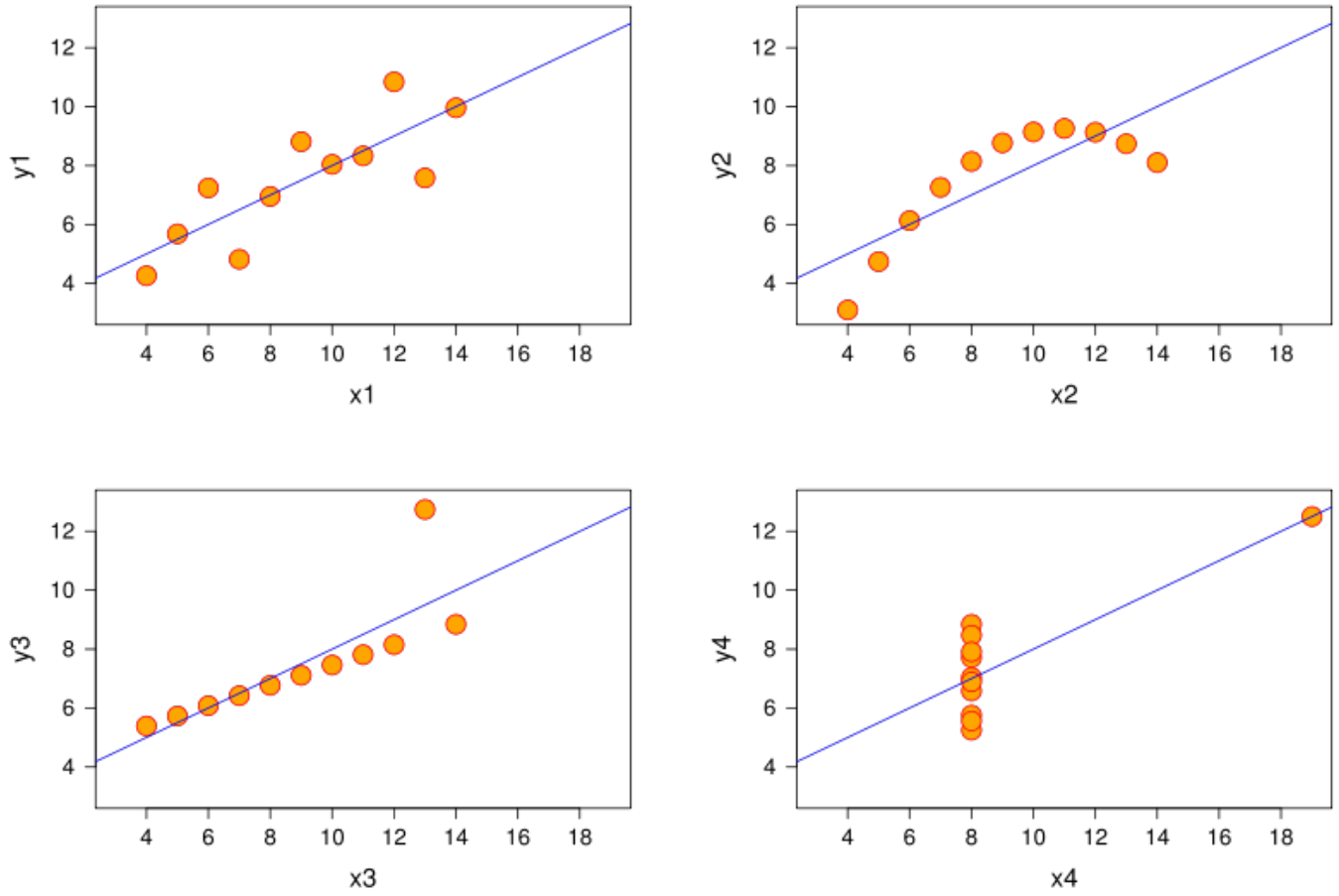
- Lots and lots of factors contribute to just about any dependent measure

Examples of correlation values



From Denis Boigelot, available on Wikipedia

Same correlation, different functions



From John Behrens, Pearson

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)

RMSE/MAD

Mean Absolute Deviation

- 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

MAD vs. RMSE

- MAD 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

Example

Actual	Pred
1	0.5
0.5	0.75
0.2	0.4
0.1	0.8
0	0.4

Example (MAD)

Actual	Pred	AD
1	0.5	abs(1-0.5)
0.5	0.75	abs(0.75-0.5)
0.2	0.4	abs(0.4-0.2)
0.1	0.8	abs(0.8-0.1)
0	0.4	abs(0.4-0)

Example (MAD)

Actual	Pred	AD
1	0.5	abs(1-0.5)=0.5
0.5	0.75	abs(0.75-0.5)=0.25
0.2	0.4	abs(0.4-0.2)=0.2
0.1	0.8	abs(0.8-0.1)=0.7
0	0.4	abs(0.4-0)=0.4

Example (MAD)

Actual	Pred	AD
1	0.5	abs(1-0.5)=0.5
0.5	0.75	abs(0.75-0.5)=0.25
0.2	0.4	abs(0.4-0.2)=0.2
0.1	0.8	abs(0.8-0.1)=0.7
0	0.4	abs(0.4-0)=0.4

MAD = avg(0.5,0.25,0.2,0.7,0.4)=0.41

Example (RMSE)

Actual	Pred	SE
1	0.5	$(1-0.5)^2$
0.5	0.75	$(0.75-0.5)^2$
0.2	0.4	$(0.4-0.2)^2$
0.1	0.8	$(0.8-0.1)^2$
0	0.4	$(0.4-0)^2$

Example (RMSE)

Actual	Pred	SE
1	0.5	0.25
0.5	0.75	0.0625
0.2	0.4	0.04
0.1	0.8	0.49
0	0.4	0.16

Example (RMSE)

Actual	Pred	SE
1	0.5	0.25
0.5	0.75	0.0625
0.2	0.4	0.04
0.1	0.8	0.49
0	0.4	0.16

MSE = Average(0.25,0.0625,0.04,0.49,0.16)

Example (RMSE)

Actual	Pred	SE
1	0.5	0.25
0.5	0.75	0.0625
0.2	0.4	0.04
0.1	0.8	0.49
0	0.4	0.16

MSE = 0.2005

Example (RMSE)

Actual	Pred	SE
1	0.5	0.25
0.5	0.75	0.0625
0.2	0.4	0.04
0.1	0.8	0.49
0	0.4	0.16

RMSE = 0.448

Note

- Low RMSE/MAD is good
- High Correlation is good

What does it mean?

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

What does it mean?

- High RMSE/MAD, 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/MAD, 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