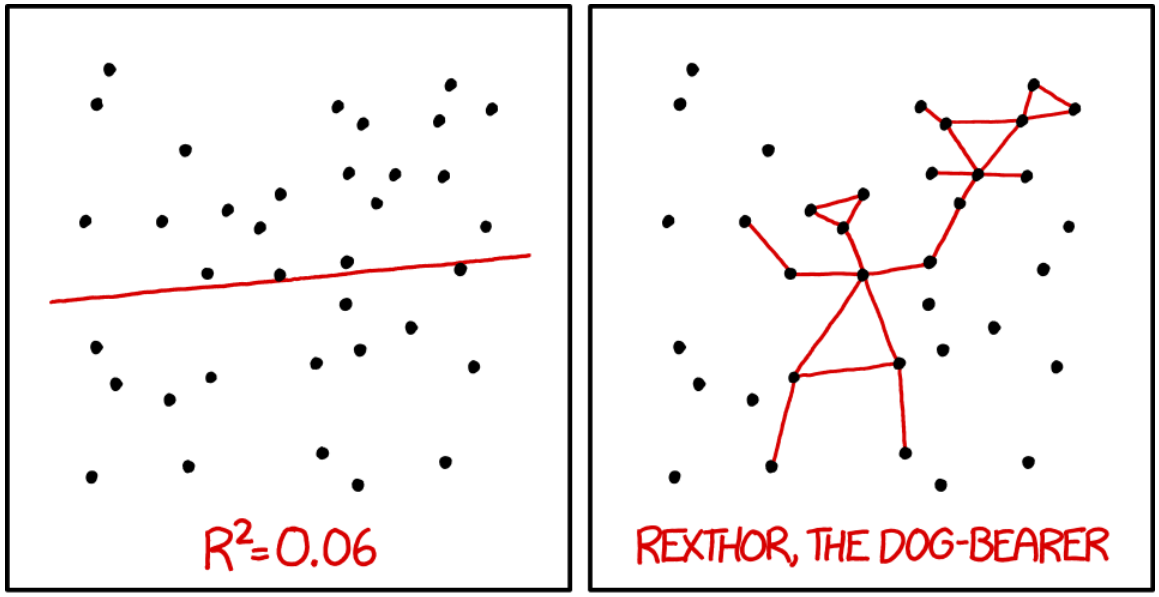
Today's question



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

How's my fit?

Decoding

Today's docket

1. Decoding

   Model types, SVM, cross-validation

2. EMG dataset

   Breakout session

In a nutshell

- Central aim is to identify an experimental stimulus category from patterns in the data
  (e.g., cats vs. dogs, werewolves vs. villagers)

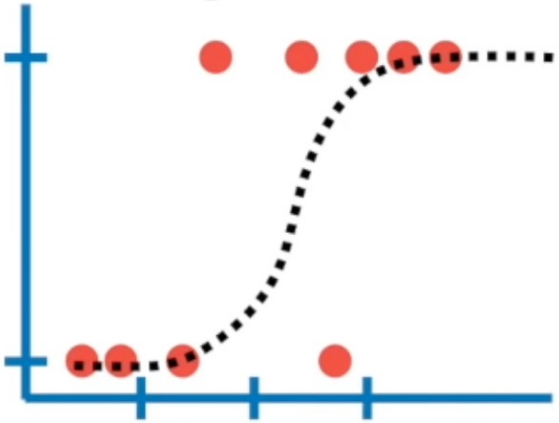
- Operationalized by fitting a *supervised learning model* on n-dimensional data ("training")
  (coordinates of points in space are usually called "features")

- With model evaluation based on generalizability to unseen or new data points ("testing")
  (i.e., "decoding accuracy")

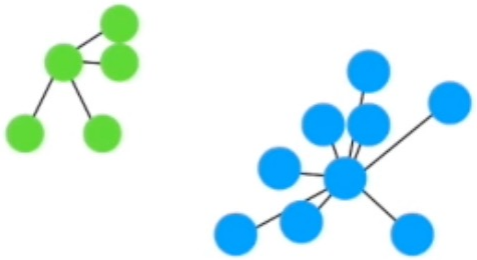
Revealing the presence of information in data

Model types

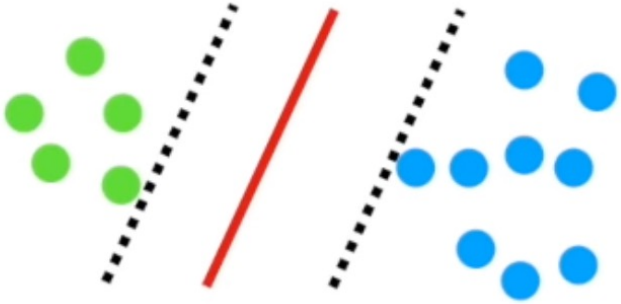
We could use Logistic Regression…



…or K-nearest neighbors…



…or support vector machines (SVM)…



HR (heart)	GSR (skin)	Talkativeness	Turn gap (ms)	Werewolf
80	17	average	190	
75	15	average	180	
93	12	average	200	
120	21	extreme	130	
70	10	average	180	

Choosing the appropriate supervised learning method

Support Vector Machines

Carving out subspaces in feature space using hyperplanes

Train vs. test set

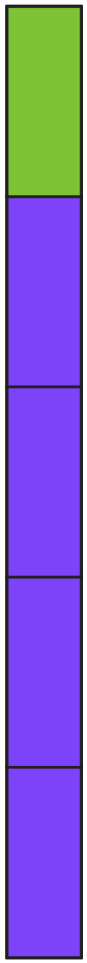
- A suboptimal approach would be to use all of the data to estimate the model parameters
  (i.e., “train” the model)

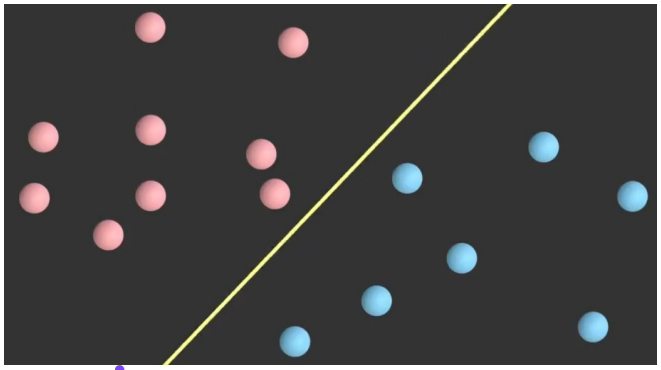
- We need to know how the model will work on data it wasn’t trained on
  (otherwise, the model is useful in reference only to its initial dataset, and not to any other datasets – “overfitting”)

- Therefore, a part of the available data needs to be held out from training
  (i.e., the “test set”)

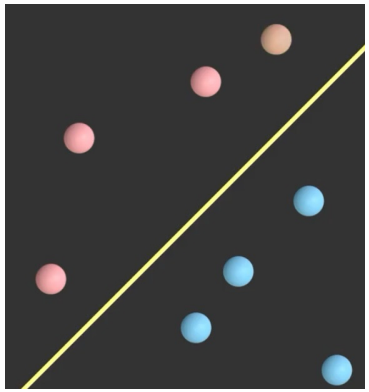
Holding out part of the available data as a test set

Cross-validation





train set



test set

78%

81%

92%

76%

83%

decoding accuracies **82%**

Rotation estimation for model evaluation

Today's docket

1. Decoding

   Model types, SVM, cross-validation

2. EMG dataset

   Breakout session

EMG dataset

- Lab8_Decoding.ipynb

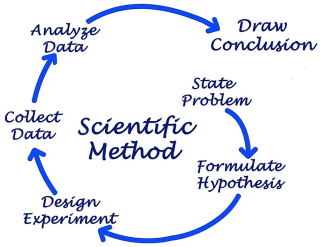
Take-home concepts

- Decoding is about revealing the presence of (categorical) information in data

Next up

- *Wrap up Data Collection asap*
- *Start Data Analysis*
- *Hackathon next week*
- *No class this Wednesday*

Planning



Week 7	Lab 7: Data Analysis Lab 8: Decoding	NO CLASS *Data Collection*
Week 8	Hackathon	**Research: Data Analysis**
Week 9	Anatomy of a Paper	**Research: Conclusion**
Week 10	11/14 *Research Report due*	