

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



1. Decoding

Model types, SVM, cross-validation

2. EMG dataset

Breakout session

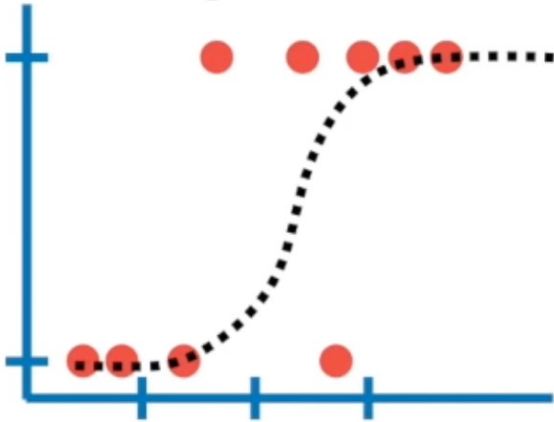
In a nutshell

- 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”)

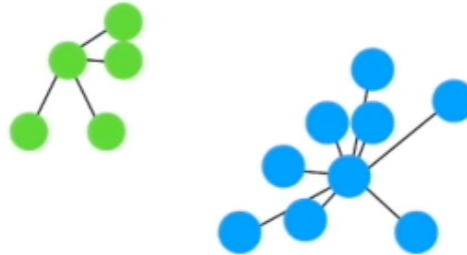
Revealing the presence of information in data

Model types

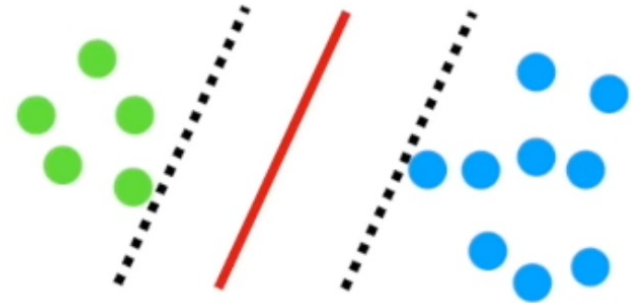
We could use Logistic Regression...



...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



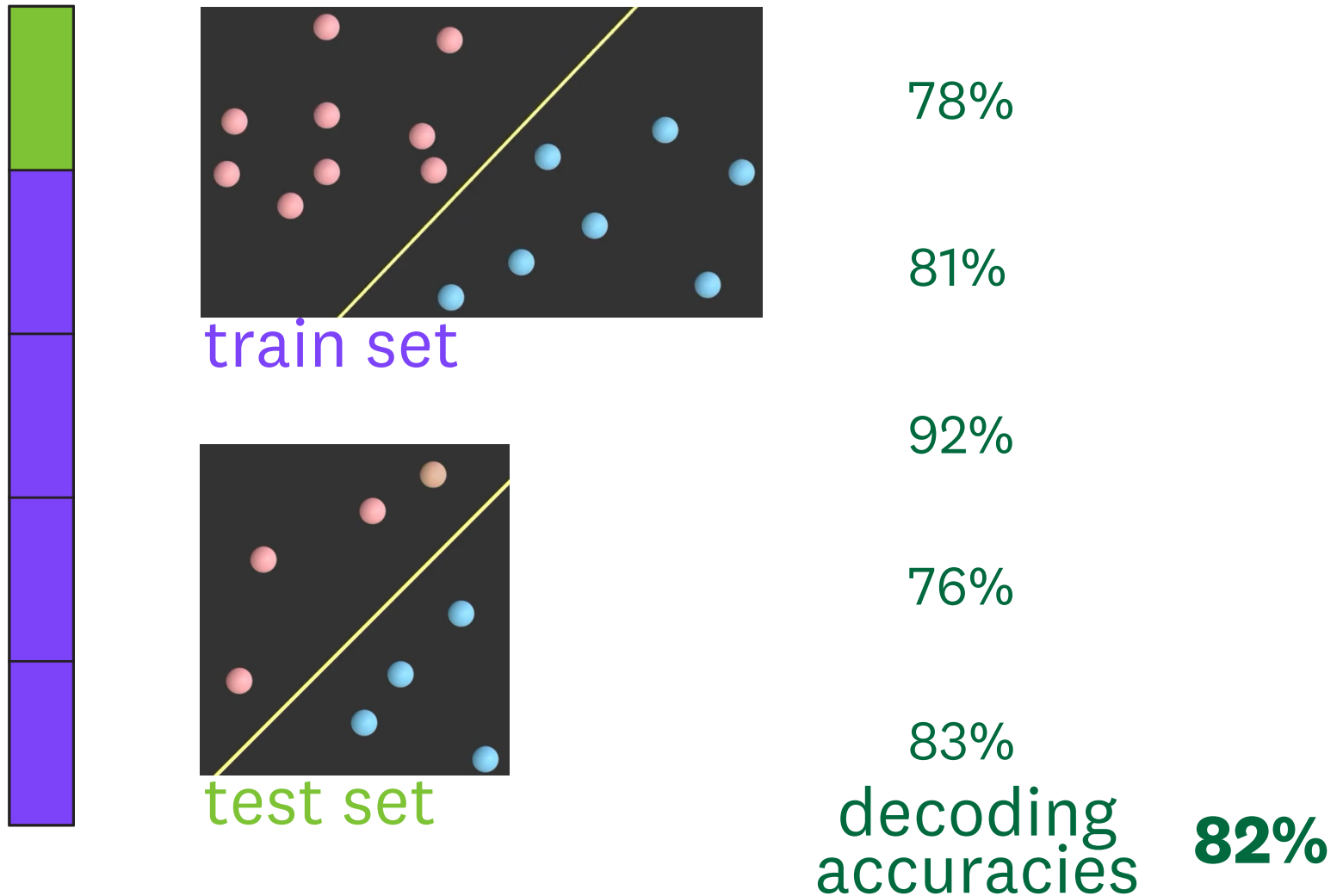
Carving out subspaces in feature space using hyperplanes

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



Transforming the raw data into an understandable format

1. Decoding

Model types, SVM, cross-validation

2. EMG dataset

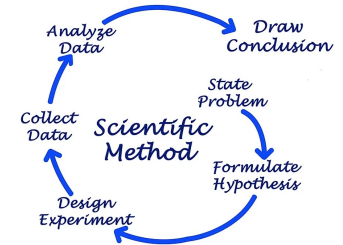
Breakout session

- Lab8_Decoding.ipynb

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

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

Planning



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