ML Method: Classification (k-Nearest Neighbors)

Presenter: Kayla Pascua
Professor: Yuen Yuen

What is machine learning?

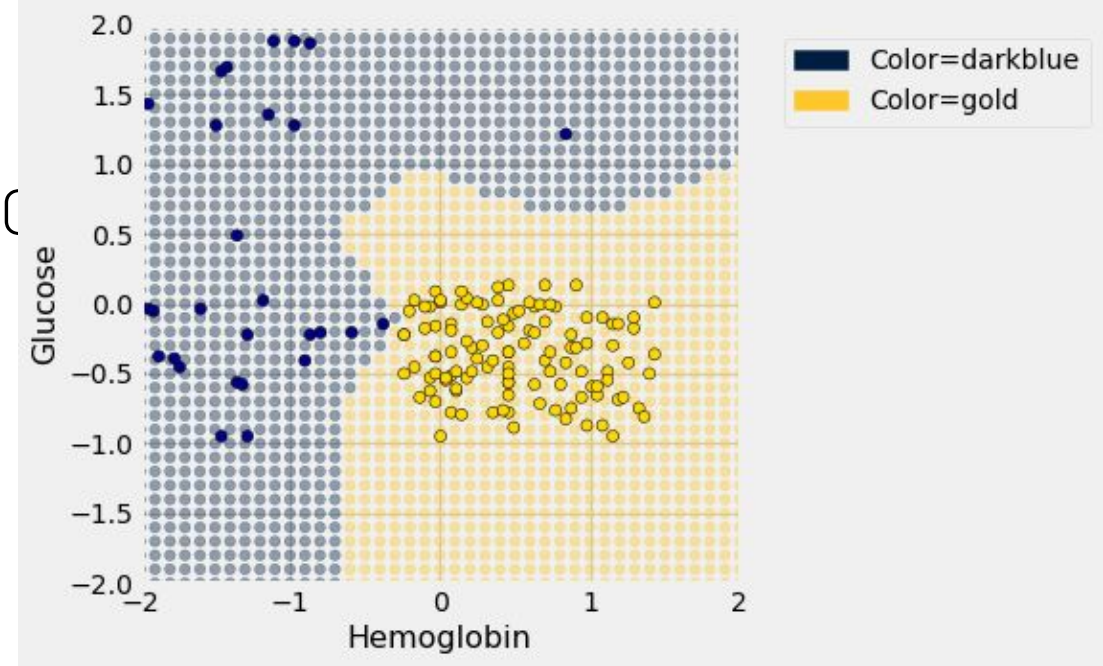
- Finding patterns in data and drawing inferences/making predictions
- Supervised, unsupervised, reinforcement

What is classification?

- Supervised learning
- When to use: Predicting categorical variables
- Types:
  - K-Nearest Neighbor
  - Support Vector Machine (SVM)
  - Logistic Regression

k-Nearest Neighbors

What does it do?



How it works

1. Training
2. Compute distance
3. Find k closest points
4. Assign class to data point

Key Parameters

- k
- Distance metric

Advantages & Disadvantages

+

- Simple
- Good with small/low dimensional datasets

-

- Computationally expensive
- Takes memory
- “Curse of dimensionality”

How to optimize

Choose Good *k* Value

- Using cross-validation

Feature Scaling

- Normalization to eliminate bias

Dimensionality Reduction

- PCA, LDA, or other techniques to handle high-dimensional data

Summary

- Simple and powerful
- Versatile and applicable in real life data
- Consideration for *k* and distance metric
- Can be optimized

Thank you!