Introduction
to
Machine learning

What is machine learning?

1. Supervised learning:
   a. Classification
   b. Regression

2. Unsupervised learning

Attribute/feature

decision/output/label

When the decision variable is continuous → regression

When the decision variable is discrete → classification

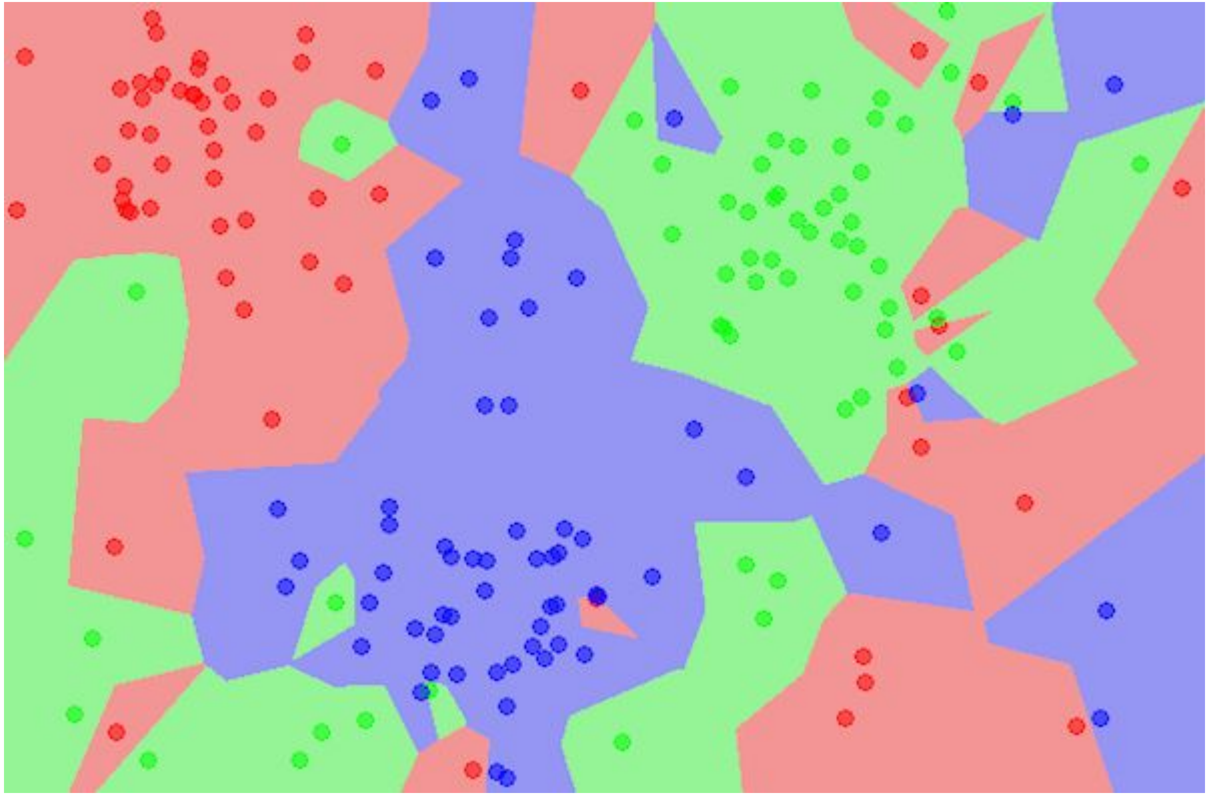
Model

Training set

Validation set

Test set

K-Nearest Neighbour



K-Nearest Neighbour

For each test input point,

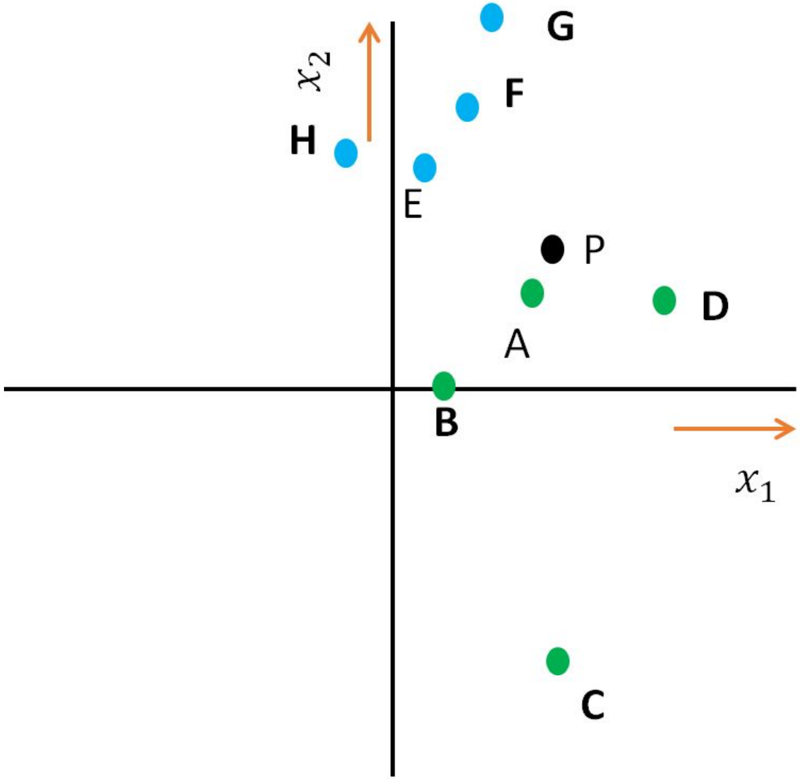
1. considers the class/output of its nearest k number of train (available) data points and
2. Determine its class by voting of the k data points

A. May use different distance calculation measure (e.g., Euclidean, Manhattan)

B. Voting system can be equal/weighted

K-Nearest Neighbour

- Calculate Distance from ***point p*** to all the training points
- If k=1, nearest point = {A} Predicted Class= Green
- If k=2, nearest points = {A, D} Predicted Class= Green
- If k=3, nearest points = {A, D, E} Predicted Class= Green
- What if k=7?



Quick question?

What did we see in the previous slide?

- Classification
- Regression

How can we tell if the model is working fine?

We need a measure.

Error or accuracy?

Error: predicted output - actual output = 15.4 - 14 = 1.4

Accuracy : how many correctly predicted labels / total test cases

Introduce datasets

- Iris
- Diabetes

KNN Algorithm

Refer to knn.txt

Deadline

30th March 11:59PM