

Machine Learning 2.1 : Classification

Théo Trouillon - theo.trouillon@le-campus-numerique.fr

Regression VS Classification

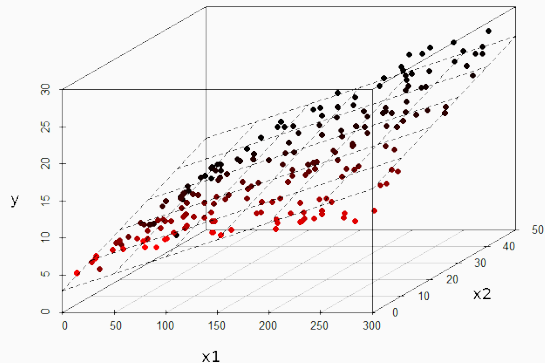
- Supervised learning
 - Feature matrix $X \in \mathbb{R}^{n \times m}$, n samples, m features
 - Target vector y of size n
 - $y \approx f(X)$

Regression VS Classification

- Supervised learning
 - Feature matrix $X \in \mathbb{R}^{n \times m}$, n samples, m features
 - Target vector y of size n
 - $y \approx f(X)$
- Regression
 - $y \in \mathbb{R}^n$
 - Linear regression :
$$y \approx w_1x_1 + w_2x_2 + \dots + w_mx_m$$

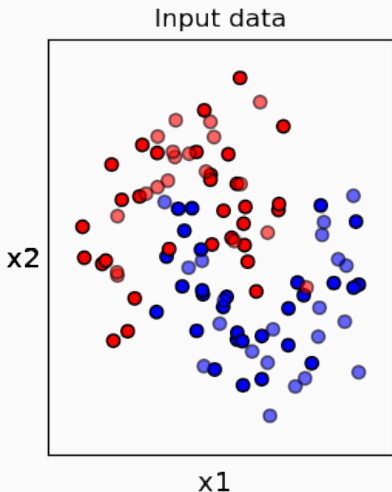
Regression VS Classification

- Supervised learning
 - Feature matrix $X \in \mathbb{R}^{n \times m}$, n samples, m features
 - Target vector y of size n
 - $y \approx f(X)$
- Regression
 - $y \in \mathbb{R}^n$
 - Linear regression :
$$y \approx w_1x_1 + w_2x_2 + \dots + w_mx_m$$

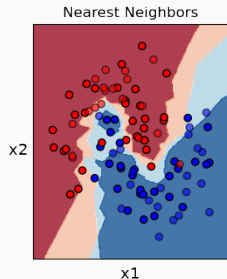
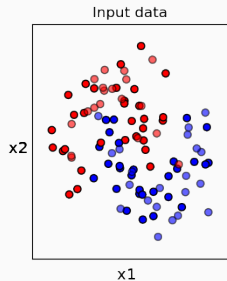


Regression VS Classification

- Supervised learning
 - Feature matrix $X \in \mathbb{R}^{n \times m}$, n samples, m features
 - Target vector y of size n
 - $y \approx f(X)$
- (Binary) Classification
 - $y \in \{0, 1\}^n$, called classes or labels
 - Red : positives, $y=1$
 - Blue : negatives, $y=0$
 - Many algorithms, today :
K-Nearest-Neighbors

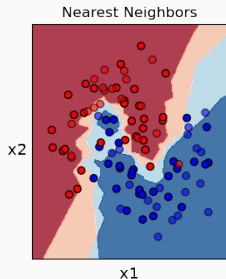
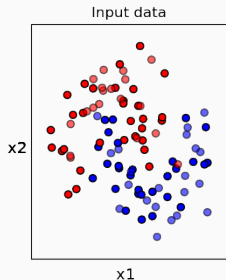


Binary classification with K-Nearest Neighbors (KNN)



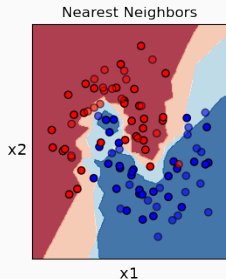
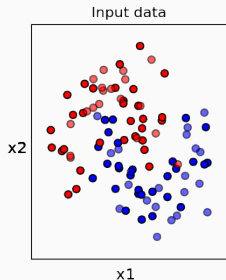
- Example of binary classification :
 - Red : positives, $y=1$
 - Blue : negatives, $y=0$

Binary classification with K-Nearest Neighbors (KNN)



- Example of binary classification :
 - Red : positives, $y=1$
 - Blue : negatives, $y=0$
- K-Nearest Neighbors :
 - Training : memorize all training points
 - Prediction : same y as the **majority** among K-nearest training points
 - Here : $K = 3$

Binary classification with K-Nearest Neighbors (KNN)



- Example of binary classification :
 - Red : positives, $y=1$
 - Blue : negatives, $y=0$
- K-Nearest Neighbors :
 - Training : memorize all training points
 - Prediction : same y as the **majority** among K-nearest training points
 - Here : $K = 3$
- Two types of error :
 - False positives (FP)
 - False negatives (FN)

Classification evaluation metrics

- Four prediction possibilities :
 - True Positives (TP): 1's correctly predicted as 1's
 - True Negatives (TN): 0's correctly predicted as 0's
 - False Negatives (FN): 1's incorrectly predicted as 0's
 - False Positives (FP): 0's incorrectly predicted as 1's

Classification evaluation metrics

- Four prediction possibilities :
 - True Positives (TP): 1's correctly predicted as 1's
 - True Negatives (TN): 0's correctly predicted as 0's
 - False Negatives (FN): 1's incorrectly predicted as 0's
 - False Positives (FP): 0's incorrectly predicted as 1's
- Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$
 - Overall classes prediction ability

Classification evaluation metrics

- Four prediction possibilities :
 - True Positives (TP): 1's correctly predicted as 1's
 - True Negatives (TN): 0's correctly predicted as 0's
 - False Negatives (FN): 1's incorrectly predicted as 0's
 - False Positives (FP): 0's incorrectly predicted as 1's
- Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$
 - Overall classes prediction ability
- Precision = $\frac{TP}{TP+FP}$
 - Ability to not predict 0's as 1's

Classification evaluation metrics

- Four prediction possibilities :
 - True Positives (TP): 1's correctly predicted as 1's
 - True Negatives (TN): 0's correctly predicted as 0's
 - False Negatives (FN): 1's incorrectly predicted as 0's
 - False Positives (FP): 0's incorrectly predicted as 1's
- Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$
 - Overall classes prediction ability
- Precision = $\frac{TP}{TP+FP}$
 - Ability to not predict 0's as 1's
- Recall = $\frac{TP}{TP+FN}$
 - Ability to not predict 1's as 0's

Today

- Heart disease diagnostic with k-nearest neighbors
 - Features : Age, cholesterol, blood pressure, ...
 - Classes : 0 : no heart disease, 1 : heart disease
 - Importance of Precision vs Recall
- Experimental setup : same as previous module : train/test, cross-validation, ...
 - Chapter 2 from *Hand's On Machine Learning* ...
- Understand the difference between regression and classification :
 - Chapter 4.1 and 4.2 from *Introduction to Statistical Learning*
- Take time to **read** the resources when mentioned in the notebook