Classification

Louis Jachiet

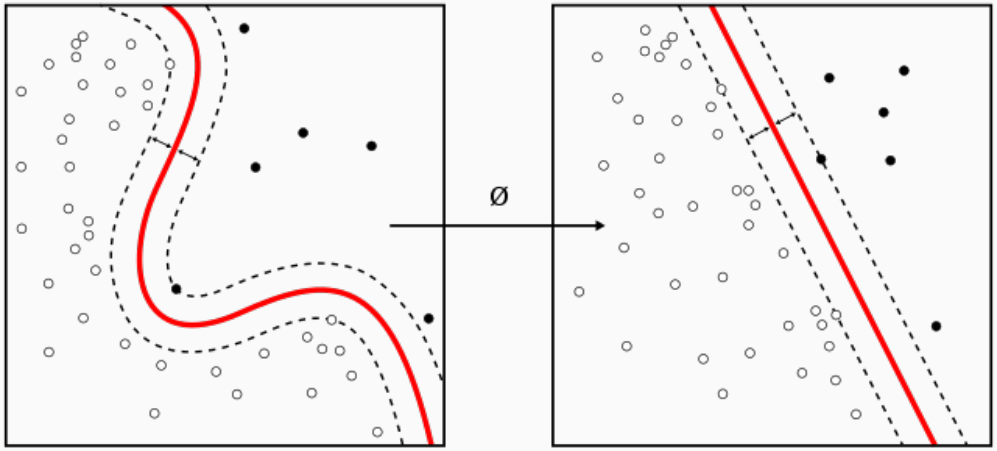
What is classification?

Wikipedia definition

Classification is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known.

Classification

Generalize known structures to apply to new data.



An e-mail program might attempt to classify an e-mail as "legitimate" or as "spam".

Spam example

Data set that describes e-mail features for deciding if it is spam.

Example

Contains "Money"	Domain type	Has attach.	Time received	spam
yes	com	yes	night	yes
yes	edu	no	night	yes
no	com	yes	night	yes
no	edu	no	day	no
no	com	no	day	no
yes	cat	no	day	yes

Spam example

Data set that describes e-mail features for deciding if it is spam.

Example

Contains "Money"	Domain type	Has attach.	Time received	spam
yes	com	yes	night	yes
yes	edu	no	night	yes
no	com	yes	night	yes
no	edu	no	day	no
no	com	no	day	no
yes	cat	no	day	yes

Assume we have to classify the following new instance:

Contains "Money"	Domain type	Has attach.	Time received	spam
yes	edu	yes	day	?

Definition

Definition

Given a set of classes $C_1 ... C_N$, a classifier algorithm builds a model that predicts for every unlabelled instance I the class C_i to which it belongs with accuracy.

Example
Spam filter

Example
Twitter Sentiment analysis: analyze tweets with positive or negative feelings

Example
Cat or Dog?

Basic Classifiers

Majority vote

Training

Compute the majority class in the dataset

Prediction

Output the majority class

k-Nearest Neighbors (k-NN)

Training

Store all instance (+ eventual index)

Prediction

Find the k closest point in the input and output the majority over those k points.

k-Nearest Neighbors (k-NN)

Training

Store all instance (+ eventual index)

Prediction

Find the k closest point in the input and output the majority over those k points.

Closest according to what metric?

L_1 *vs* L_2 *vs* L_∞ *vs* *COS*

Bayes theorem

Formula

$$\frac{P(A) \times P(B|A)}{P(B)} = P(A|B)$$

Bayes theorem

Proof.

$$P(A \cap B) = P(A) \times P(B|A)$$

Bayes theorem

Proof.

$$P(A \cap B) = P(A) \times P(B|A)$$

$$P(A \cap B) = P(B) \times P(A|B)$$

Bayes theorem

Proof.

$$P(A \cap B) = P(A) \times P(B|A)$$

$$P(A \cap B) = P(B) \times P(A|B)$$

$$P(A) \times P(B|A) = P(B) \times P(A|B)$$

Bayes theorem

Proof.

$$P(A \cap B) = P(A) \times P(B|A)$$

$$P(A \cap B) = P(B) \times P(A|B)$$

$$P(A) \times P(B|A) = P(B) \times P(A|B)$$

$$\frac{P(A) \times P(B|A)}{P(B)} = P(A|B)$$

□

Bayes theorem

Formula

$$\frac{P(A) \times P(B|A)}{P(B)} = P(A|B)$$

Interpretation

$$\text{prior} \times \frac{\text{likelihood}}{\text{evidence}} = \text{posterior}$$

Naive Bayes Classifier

Grouping attributes

$$P(C_i) \times \frac{P(\bar{x}|C_i)}{P(\bar{x})} = P(C_i|\bar{x})$$

Multiple attributes

$$P(C_i) \times \frac{\prod_j P(x_j|C_i)}{P(\bar{x})} = P(C_i|\bar{x})$$

Naive Bayes Classifier

Grouping attributes

$$P(C_i) \times \frac{P(\bar{x}|C_i)}{P(\bar{x})} = P(C_i|\bar{x})$$

Multiple attributes

$$P(C_i) \times \frac{\prod_j P(x_j|C_i)}{P(\bar{x})} = P(C_i|\bar{x})$$

With independence hypothesis!

Naive Bayes Classifier

Grouping attributes

$$P(C_i) \times \frac{P(\bar{x}|C_i)}{P(\bar{x})} = P(C_i|\bar{x})$$

Multiple attributes

$$P(C_i) \times \frac{\prod_j P(x_j|C_i)}{P(\bar{x})} = P(C_i|\bar{x})$$

With independence hypothesis!

$P(\bar{x})$ does not change with the class

Tree Methods

Classification

Data set that describes e-mail features for deciding if it is spam.

Example

Contains "Money"	Domain type	Has attach.	Time received	spam
yes	com	yes	night	yes
yes	edu	no	night	yes
no	com	yes	night	yes
no	edu	no	day	no
no	com	no	day	no
yes	cat	no	day	yes

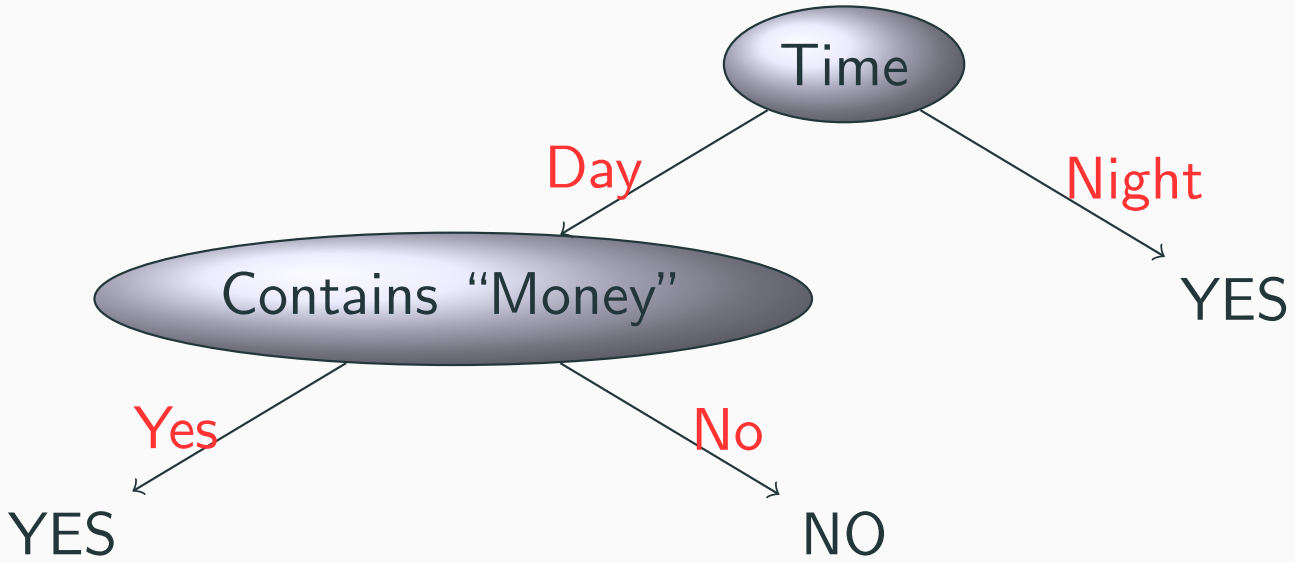
Assume we have to classify the following new instance:

Contains "Money"	Domain type	Has attach.	Time received	spam
yes	edu	yes	day	?

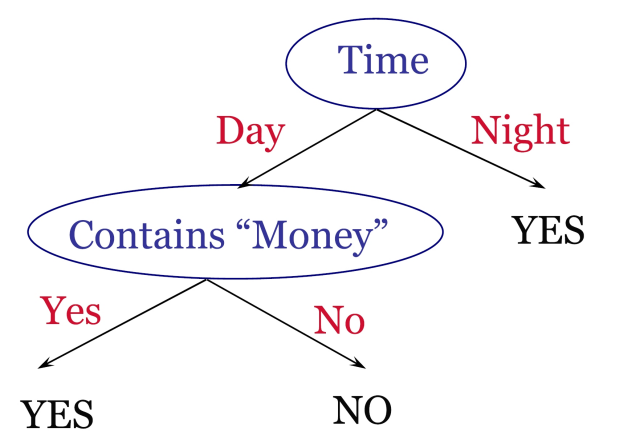
Classification

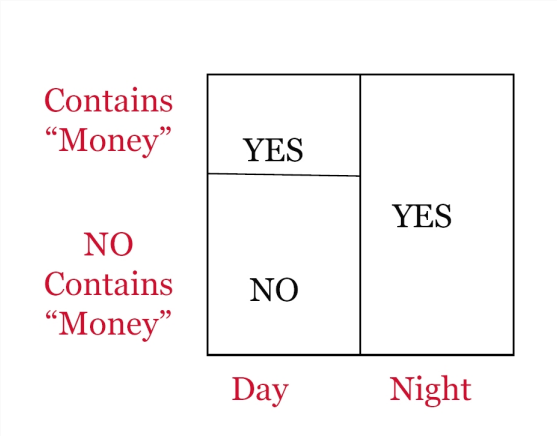
- Assume we have to classify the following new instance:

Contains "Money"	Domain type	Has attach.	Time received	spam
yes	edu	yes	day	?



Decision Trees





Recursive construction technique

- $A \leftarrow$ the *best* decision attribute for next *node*
- Assign A as decision attribute for *node*
- For each value of A, create new descendant of *node*
- Sort training examples to leaf nodes
- If training examples perfectly classified, Then STOP, Else iterate over new leaf nodes

Bagging

Example

Dataset of 4 Instances : A, B, C, D

Classifier 1: B, A, C, B

Classifier 2: D, B, A, D

Classifier 3: B, A, C, B

Classifier 4: B, C, B, B

Classifier 5: D, C, A, C

Bagging builds a set of M base models, with a bootstrap sample created by drawing random samples with replacement.

Random Forests

- Bagging
- Random Trees: trees that in each node only uses a random subset of the attributes

$\Rightarrow$ one of the most popular methods in machine learning.

Gradient-based Methods

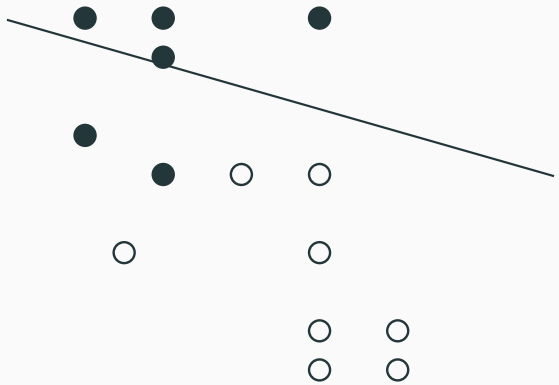
Logistic Regression

Training

Learn an hyperplan $\mathcal{P}$ separating well the two classes.

Prediction

What side of the hyperplan $\mathcal{P}$ is the point?



Based on the gradient of the logit function.

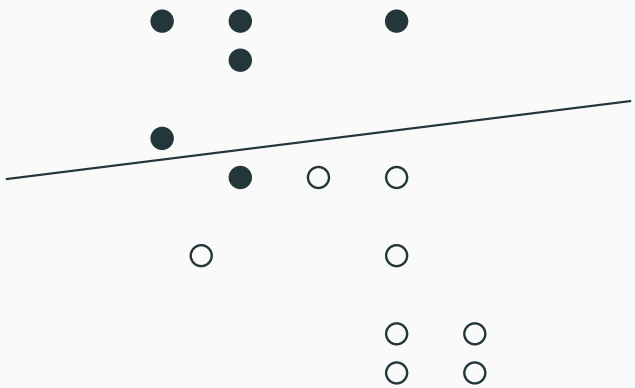
Logistic Regression

Training

Learn an hyperplan $\mathcal{P}$ separating well the two classes.

Prediction

What side of the hyperplan $\mathcal{P}$ is the point?



Based on the gradient of the logit function.

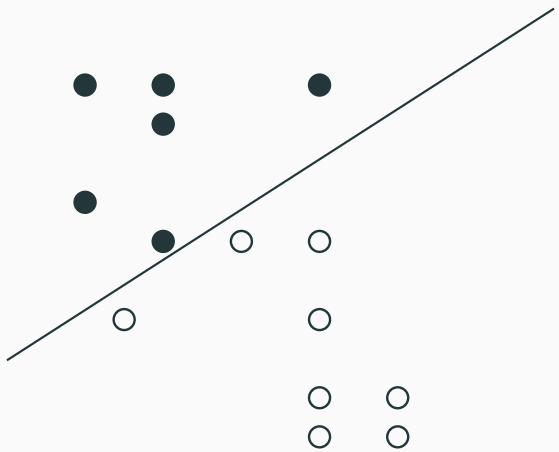
Logistic Regression

Training

Learn an hyperplan $\mathcal{P}$ separating well the two classes.

Prediction

What side of the hyperplan $\mathcal{P}$ is the point?



Based on the gradient of the logit function.

Neural Network