Aula 1 - Logistics

Aula 2 Introdução

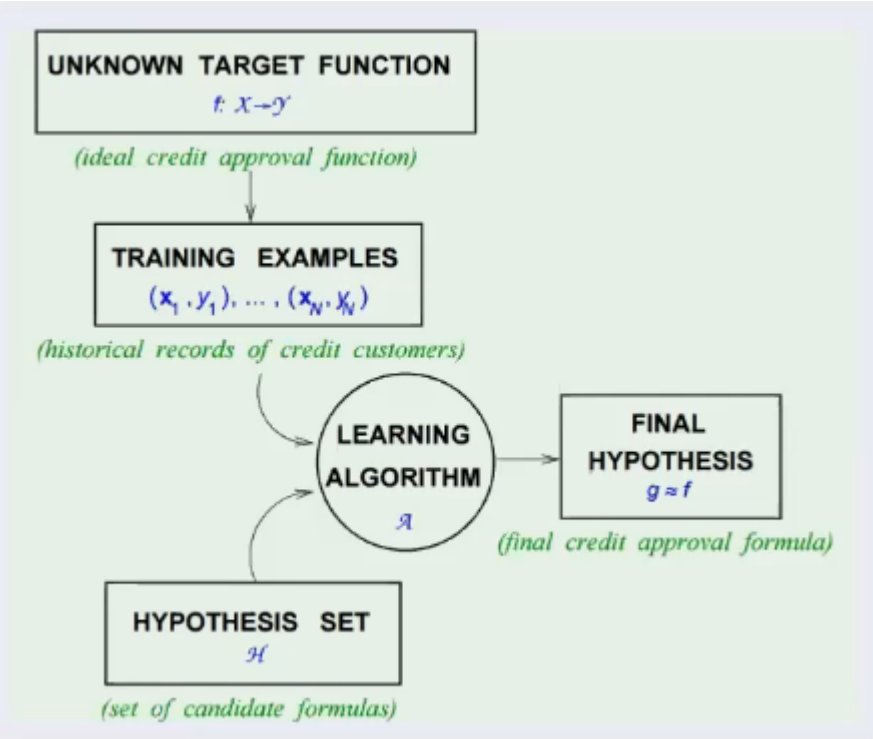
- Training set: a set of $(X, y)^m$ pairs, where input $X \in R^d$ and output $y \in \{0, 1\}$
- Goal: Learn function/model $f : X \rightarrow y$ to predict correctly on new inputs X
  - Step 1: Choose a learning algorithm
    - logistic regression, SVMs, KNNs, decision trees, ANNs etc.
  - Step 2: Optimize parameters/weights W using the training set
    - Minimize a loss function: $\min_W \sum_{i=1}^{m} (f_w(x^{(i)}) - y^{(i)})^2$

Treinamento: tunning dos parametros utiliando algoritmos de otimização para minimizar função de perdas.

Generalização: Similarmente a navalha de occam, utilizar hipoteses e funções suficientemente complexas, ou o mais simples possivel que represente os dados.

- Input: X (customer application)
- Output: y (good/bad customer?)
- Target function: $f : \mathcal{X} \rightarrow \mathcal{Y}$ (ideal credit approval formula)
- Data: $(X_1, y_1), (X_2, y_2), \ldots, (X_N, y_N)$ (training set)
- Hypothesis: $g : \mathcal{X} \rightarrow \mathcal{Y}$ (g approximates f well)

- We want to approximate a target function.
- Need a sample of data generated from the target function.
- A hypothesis set is used to avoid candidate functions that do not approximate well the target function.



O algoritmo de aprendizado procura por uma boa função no grupo de hipoteses (seja otimização ou heuristica)

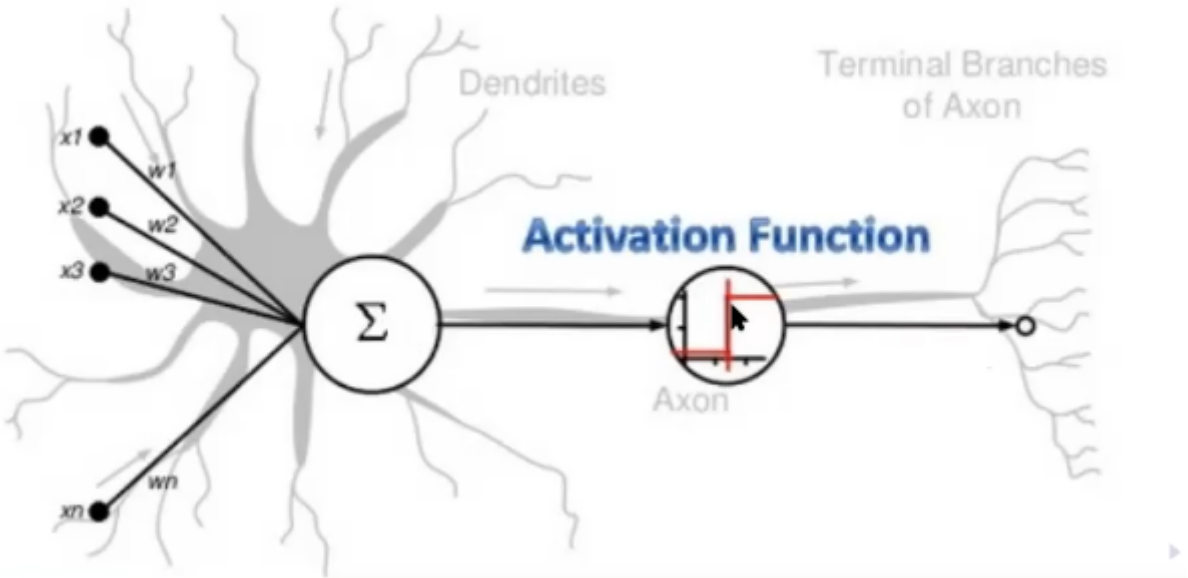
Perceptron

Simplificação de neurônio:

Multipas entradas e uma saida Cada entrada tem um peso que é multiplicado pelas entradas O neuronio faz a soma ponderada das entradas de acordo com os pesos para determinar a saida de acordo com uma funcao de ativacao

Credit approval with a linear classifier.

- Given an input $X = (x_1, x_2, \ldots, x_d)$:
  - Approve credit if $\sum_{i=1}^{d} w_i \times x_i >$ threshold,
  - Deny credit if $\sum_{i=1}^{d} w_i \times x_i <$ threshold.
- This rule can be written as:
  - $h(X) = sign\left(\left(\sum_{i=1}^{d} w_i \times x_i\right) - \text{threshold}\right)$
- Searching for h is to find optimum values for w_i and threshold
  - w_i is high if x_i is evidence for approval.
  - w_i is low if x_i is evidence for denial.
  - $W^T X = \sum_{i=1}^{d} w_i \times x_i = 0.$



Aula 3 Introdução

Credit approval with a linear classifier.

- $h(X) = sign\left(\left(\sum_{i=1}^{d} w_i \times x_i\right) - \text{threshold}\right)$
- $h(X) = sign\left(\left(\sum_{i=1}^{d} w_i \times x_i\right) + w_0\right)$
  - Threshold is $-w_0$.
- Introduce an artificial coordinate x_0 which is always set to 1.
  - $h(X) = sign\left(\sum_{i=0}^{d} w_i \times x_i\right)$
- In vector form: $h(X) = sign(W^T X)$
  - This is the hypothesis set (e.g., linear separations).