Review of Linear Regression, Logistic Regression, and K-Nearest Neighbors

Lesson 3 – Section 3

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Quick Recap

- Mathematical Framework of Machine Learning.
- How Machine Learns from Training Data?
  - Stochastic Gradient Descent
- Three Common Pitfalls in Machine Learning
  - Overfitting
  - Target leaker
  - Models not applicable where features by the model is not available when the prediction needs to be made

Overview

Review of linear regression (regression)

Review of logistic regression (classification)

Quick description of K-Nearest Regression (regression/classification)

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Linear Regression (Review)

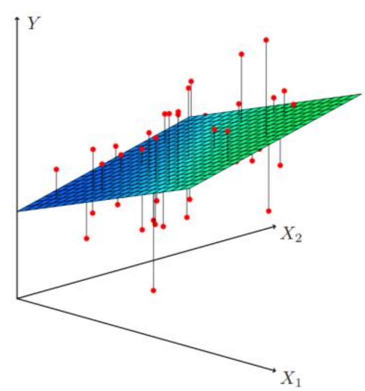
- Mathematical Equation

$$y_i = \boldsymbol{X}_i\boldsymbol{\theta} + \varepsilon_i, i = 1, 2, 3, \ldots N$$

$$\boldsymbol{\beta}=[\beta_0, \beta_1, \ldots, \beta_d]^T$$

$$\boldsymbol{X}_i=[1, x_1^{(1)}, x_1^{(2)}, \ldots, x_1^{(d)}]^T$$

$$\varepsilon \sim N(0, \sigma^2)$$



- Close form solution:

$$\boldsymbol{\theta} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}\boldsymbol{X}^\intercal \boldsymbol{y}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1^{(1)} & \ldots & x_d^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(i)} & \ldots & x_d^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & \ldots & x_d^{(n)} \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

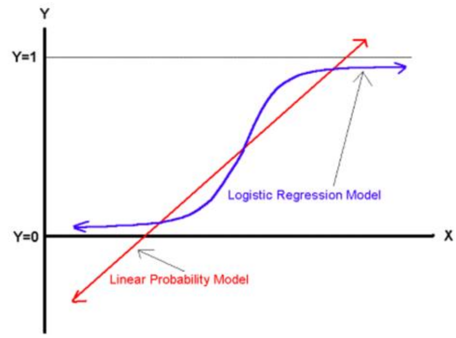
Logistic Regression

- Logistic regression (Review):

$p = Prob(Y = 1)$

$\log(\frac{p_i}{1-pi}) = \boldsymbol{X}_i\boldsymbol{\beta} + \varepsilon_i$

$$p_i = \frac{e^{\boldsymbol{\beta X}_i}}{1 + e^{\boldsymbol{\beta X}_i}}$$



- Multiclass Logistic Regression: one vs others

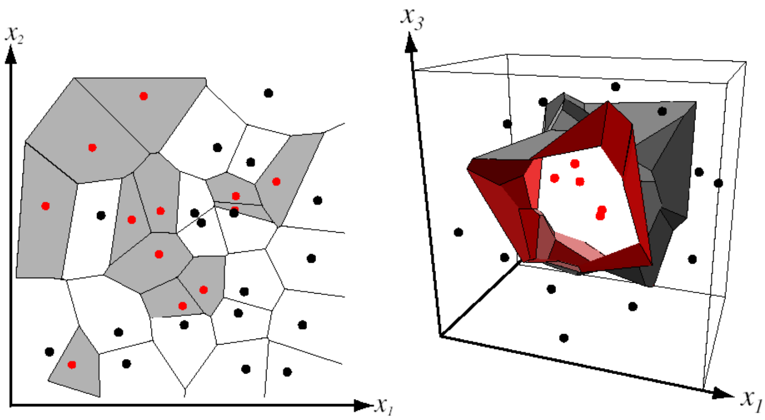
$$p(y = c \mid \boldsymbol{x}; \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_C) = \frac{\exp(\boldsymbol{\theta}_c^\top \boldsymbol{x})}{\sum_{c=1}^{C} \exp(\boldsymbol{\theta}_c^\top \boldsymbol{x})}$$

– Called the **softmax** function

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Nearest Neighbor Classifier

- Assign label of nearest training data point to each test data point

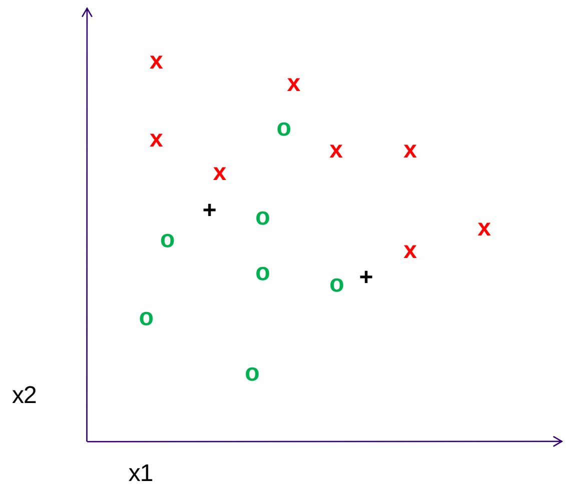


Voronoi partitioning of feature space for two-category 2D and 3D data

from Duda *et al.*

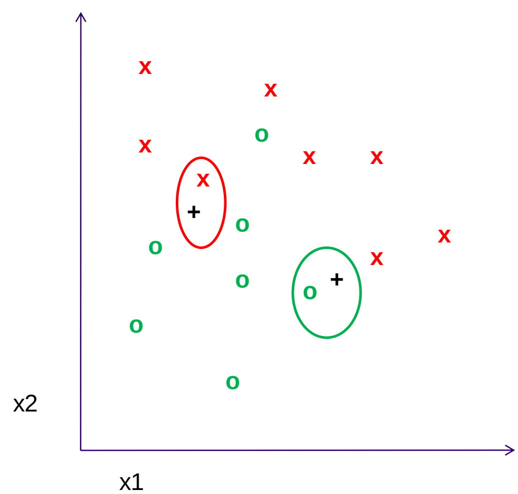
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

K-nearest neighbor



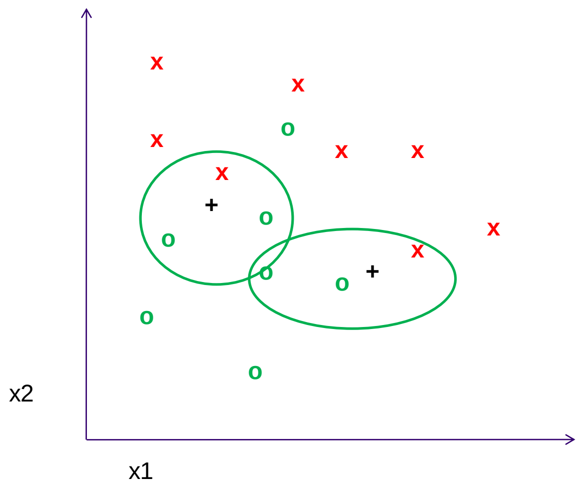
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

1-nearest neighbor



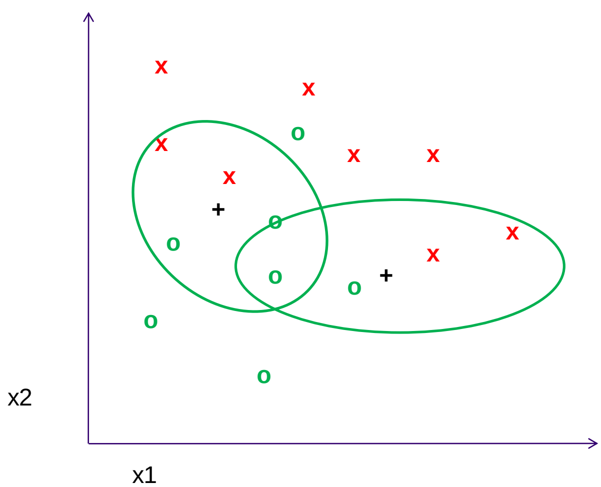
PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

3-nearest neighbor



PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

5-nearest neighbor



PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Using K-NN

- Simple, an easy one to implement
- Computationally intensive:
  - For N observations in training data, M observations in validation data, compute complexity is N*M.
- Sensitive to the choice of K and the similarity measurement
  - If K is small, very sensitive to noise in training data
  - If K is large, predicted value tends to be the mean of the training mean. Not helpful.
  - Similarity measurement matters as well.

PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Summary

>Linear regression (review)

>Logistic regression (review)

>K-nearest neighbors and its pros and cons.