Homework 1 Basic Machine Learning

For the deadline see Canvas

Version: Mon 3$^{\text{rd}}$ Sept, 2018 at 09:31.

Introduction

1. This is a group assignment, so sign up in groups of two students.
2. It gives you an indication of the Math used in the Basic Machine Learning course. If you don't know some terminology used or the meaning of some of the equations then please catch up your Math.
3. Each group has to submit a **pdf** with their answers and explanation. **Please put your names and group at the top of the hand in**.
4. For questions about this homework assignment use the Discussion Board on Canvas.
5. Of course you may use a calculator or a programming environment such as Matlab or Python. **But your report should not contain any code. Explain your computations and results in English!**

Exercise 1: Some geometry

Part a

Given the points $(3, 1)$ and $(-1, 2)$. Give the equation of the line through these two points. The equation should have the form

$$w_0 + w_1 x_1 + w_2 x_2 = 0$$

In this equation (w_0, w_1, w_2) are the so-called weights of the line.

Part b

Make a plot of the line and the vector (w_1, w_2). The direction of (w_1, w_2) is called the positive side of the line.

Part c

Consider the point $(x_1, x_2) = (4, 3)$. Is this point on the positive side or negative side of the line? Also compute $w_0 + w_1 x_1 + w_2 x_2$ for this point.

Part d

Show that the distance of the point $(x_1, x_2) = (4, 3)$ to the line is equal to

$$|w_0 + w_1 x_1 + w_2 x_2| / \|(w_1, w_2)\|$$

where $\|(w_1, w_2)\|$ is the length of the vector (w_1, w_2).

Part e

What is the distance of the line to the origin $(0, 0)$ and how can this be computed in terms of (w_0, w_1, w_2)?

Exercise 2: Some calculus

In this exercise we consider the logistic function σ defined by

$$\sigma(y) = 1/(1 + e^{-y})$$

Part a

Show that the derivative of σ with respect to y, i.e. $\sigma'(y)$, is given by

$$\sigma'(y) = \sigma(y)(1 - \sigma(y))$$

Part b

Show that:

- for y large $\sigma(y) \approx 1$.
- for y large $\sigma'(y) \approx 0$

Part c

Now let's combine this with exercise 1: assume that $y = w_0 + w_1 x_1 + w_2 x_2$. Show that the partial derivative of σ with respect to x_1 is given by

$$\frac{\partial \sigma}{\partial x_1}(y) = \sigma(y)(1 - \sigma(y))w_1$$

Hint: use the chain rule and part a.

Part d

Once again assume $y = w_0 + w_1 x_1 + w_2 x_2$. Show that:

- $\frac{\partial \sigma}{\partial w_1}(y) = \sigma(y)(1 - \sigma(y))x_1$
- the gradient of σ with respect to the vector $w = (w_0, w_1, w_2)$ (denoted by $\nabla_w \sigma(y)$) is given by $\nabla_w \sigma(y) = \sigma(y)(1 - \sigma(y))x$, where x is the vector $x = (1, x_1, x_2)$

Part e

Once again assume $y = w_0 + w_1 x_1 + w_2 x_2$. Show that:

- if x_1 is positive then for w_1 large $\frac{\partial \sigma}{\partial w_1}(y) \approx 0$.
- if x_1 is negative then for w_1 large $\frac{\partial \sigma}{\partial w_1}(y) \approx 0$.

Exercise 3: Some statistics

Consider the very small dataset $X = \{(1, 3), (3, 5), (-2, 1), (-2, -3)\}$.

Part a

Compute the mean of X.

Part b

Compute the covariance matrix of X. This should be a 2x2 matrix.

Part c

Compute, based on the covariance matrix of part b, the correlation between x_1 (the first component) and x_2.

Part d

Compute the eigenvalues and eigenvectors of the covariance matrix of part b. What would be the covariance matrix if we transform the data points into coordinates with respect to these eigenvectors?