Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNIA Winter Term 2018/2019

Exercise Sheet 4

Machine Learning Basics

Deadline: 26.11.2018, 23:59

Exercise 4.1 - Gradient Descent and Newton's Method (1.5 + 1.5 + 0.5 + 0.5 = 4 points)

In the optimization setting, Gradient Descent and Newton's Method are two commonly used iterative methods for finding a solution. Let's say we want to minimize a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ which is defined as $f(\mathbf{x}) = x_1^2 - 3x_1 + x_2^2 - x_1x_2$. The starting point is $\mathbf{x}^{(0)} = [1, 1]^T$.

a) Use Gradient Descent with step size (also known as learning rate) $\epsilon = 0.5$ to minimize the function f. The iteration should stop if the L2-norm of the gradient at the current point is less than 0.2. Show your intermediate steps. [Hint: The iteration should finish within 4 steps].

b) Starting from the same point $\mathbf{x}$, use Newton's Method to find a solution. Is the solution you get a global minimum? Argue why or why not.

c) We learned from the previous Exercise 3.3 that a bad step size can lead to non-local minimum for Gradient Descent. But are we guaranteed to converge to a local minimum when using Newton's Method since it does not have an explicit step size? What is the implicit step size used in Newton's Method?

d) Is Newton's Method always applicable if the function f is twice continuously differentiable? Argue with the help of the function $f(x) = 2x^3 - 5x$ at $x = 0$.

Exercise 4.2 - Overfitting (0.5 + 1.5 + 1 = 3 points)

Overfitting happens when the model capacity is too high and there is no proper regularization applied. Figure 1 and Figure 2 show two different classification boundaries for a binary classification problem, where the blue points and the red points represent the training data consisting out of two classes. Please answer the following questions.

a) Which classification boundary correspond to the overfitting and the underfitting, respectively?

b) Explain the terms overfitting, underfitting, and model capacity with the help of Figure 1 and Figure 2.

c) What happens to the training error and validation error when a model overfits? Explain.

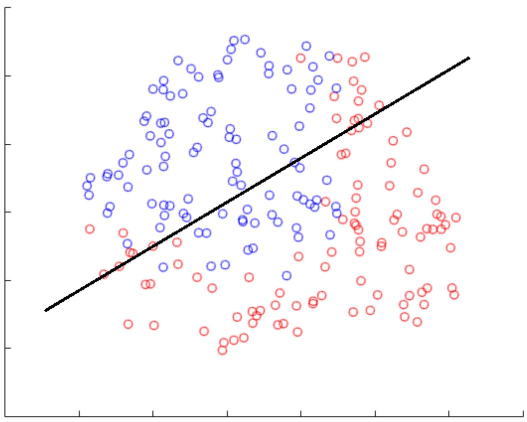

Figure 1

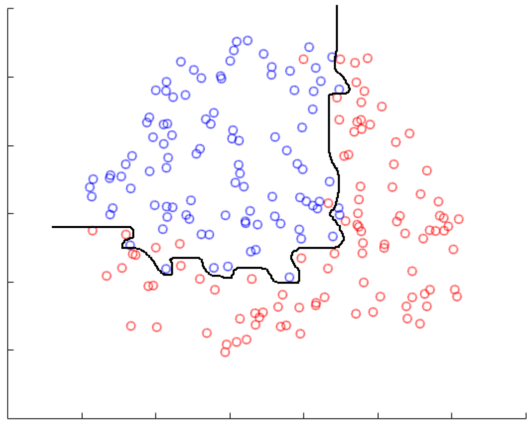

Figure 2

Exercise 4.3 - Regularization (1 + 1 + 1 = 3 points)

The solution of the linear regression problem given on slide 14 of chap5 is also known as *normal equation*. However, the inverse of $\mathbf{x}^T\mathbf{x}$, where $\mathbf{x}$ is the design matrix, may not exist. Therefore you may have to solve a linear system rather than applying the closed form solution directly. We saw in the lecture a modified linear regression problem, where we minimize a loss function $J(\mathbf{w}) = MSE_{train} + \lambda\mathbf{w}^T\mathbf{w}$ for $\mathbf{w}$, $\lambda > 0$. This modified version is known as *ridge regression*. One of the purposes of ridge regression is to obtain a unique solution for the normal equation when $\mathbf{x}^T\mathbf{x}$ is not invertible.

a) Given the same setting of linear regression from the lecture (see slide 14, chap5), derive the closed form solution for the ridge regression. [Hint: you may want to use some formulas from Exercise 3.1.]

b) Argue (no need for rigorous proof) for the uniqueness of the closed form solution for the ridge regression, i.e. argue why the linear system has a unique solution.

c) Argue (no need for rigorous proof) that the solution you get from a) is indeed a global minimizer.

Submission instructions

The following instructions are mandatory. If you are not following them, tutors can decide to not correct your exercise.

- You **have to** submit the solutions of this assignment sheet as a team of 2-3 students.
- Hand in a **single** PDF file with your solutions.
- Therefore make sure to write the student ID and the name of each member of your team on your submission.
- Your assignment solution must be uploaded by only **one** of your team members to the course website.
- If you have any trouble with the submission, contact your tutor **before** the deadline.
- **Don't forget to attach a solution of exercise 3.2 (if not already submitted).**