Assignment 3

Machine Learning: Algorithms and Theory

Prof. Ulrike von Luxburg / Diego Fioravanti / Moritz Haas / Tobias Frangen

Summer term 2018 — due to **May 8th**

Exercise 1 (Multidimensional derivatives, 1 + 2 points)

(a) Let $f(X) = a^T X$, where $X \in \mathbb{R}^3$ is a column vector, and $a^T = [2, -1, 5]$. Compute the following derivative $\frac{\partial f}{\partial X}$.

(b) Let $f(X) = X^T A X$, where $X \in \mathbb{R}^2$ is a column vector of two elements and

$$A = \begin{pmatrix} 1 & 2 \\ 5 & 3 \end{pmatrix}$$

Compute the following derivative: $\frac{\partial f}{\partial X}$.

You may check if your results match with the derivation rules of the following table : `https://en.wikipedia.org/wiki/Matrix_calculus#Scalar-by-vector_identities`. **Note:** You are not allowed to just use the table to come up with the solutions.

Exercise 2 (Optimality and uniqueness for linear regression, 2+4+1 points)

In this exercise let $X \in \mathbb{R}^{n \times d}$ be a matrix.

(a) Show that $X^t X$ is positive semidefinite and that $\operatorname{rank}(X^t X) = \operatorname{rank}(X)$.

(b) Suppose $\operatorname{rank}(X^t X) \leq d$ and let ω^* be an optimal solution for the linear regression problem

$$\min_{\omega} \|Y - X\omega\|^2 \tag{1}$$

show that $X(\omega^* - \omega_0) = 0$ where $\omega_0 = (X^t X)^+ X^t Y$.

(c) Use part (b) to prove that if $\operatorname{rank}(X^t X) = d$ then (1) has an unique solution.

Exercise 3 (Manual linear and ridge regression, 2+3+2 points)

Let

$$X_1 = (1,1,0) \qquad X_2 = (1,-1,2) \qquad X_3 = (2,3,-1) \qquad X_4 = (-1,2,-3)$$
$$Y_1 = 3 \qquad Y_2 = 1 \qquad Y_3 = 7 \qquad Y_4 = 0$$

(a) Find 3 optimal solutions for the linear regression problem

$$\min_{\omega} \|Y - X\omega\|^2$$

(b) Predict $\hat{Y}$ for $X_5 = (0, 2, -2)$ and $X_6 = (1, 0, 0)$ using each of the 3 solutions that you found in (a). Can you explain why the prediction matches in one case but not in the other?

(c) Find 3 optimal solutions for the ridge regression problem

$$\min_{\omega} \|Y - X\omega\|^2 + \lambda\|\omega\|^2 \text{ where } \lambda = 1$$

Exercise 4 (Linear and ridge regression in python, 1+2+1+3+1+2 points)

In this exercise you will implement linear and ridge regression.

(a) Let

$$x \sim \text{Unif}([0,2])$$
$$y(x) \sim 2\sin 2x + \varepsilon$$

Where $\varepsilon \sim \mathcal{N}(0,2)$. Do a scatter plot of the sampled points.

(b) Write a function

```
def ridge_regression(X, Y, lam=1)
```

that, given an $n \times D$ matrix `X` an and $X \times 1$ vectors `Y`, returns the weights ω of computed by the ridge regression.

(c) For λ use the values $0.1, 1, 10$, compute the the ridge regression for the values sampled as in a) and compare the mean squared error (MSE). Plot all three predictions.

(d) For every x as in a) compute the representation of x in the basis given by

$$\{1, x, x^2\}$$

and then perform ridge regression on this basis. Use $\lambda = 0.001, 0.01, 0.1, 1, 10$ and compare the MSE. Plot the best prediction. Set $\lambda = 0$, see what happens with the linear regression and explain what you observe.

(e) In `X_test` and `Y_test` you will find 20 new samples. Run linear and ridge regression (trained on the original samples) on them for $\lambda = 0.001, 0.01, 0.1$. Which one works best and why?

(f) Let

$$x \sim \text{Unif}([0,2]^2)$$
$$y(x) \sim 2x_1^2 + 2x_2 + 1 + \varepsilon$$

where $\varepsilon \sim \mathcal{N}(0,2)$. For every x compute the representation of x in the basis given by

$$\{1, x_1, x_1^2, x_2, x_2^2, x_1x_2\}$$

For $\lambda = 0.001, 0.01, 0.1, 1, 10$ perform ridge regression on this basis, using x and y as inputs, and plot the MSE as a function of λ.