Universität des Saarlandes
FR 4.7 Sprach- und Signalverarbeitung

Prof. Dietrich Klakow
Pattern and Speech Recognition

Exercise Sheet 3

Machine Learning Basics

Deadline: 29.11.2016, 23:59

Linear Regression

Exercise 3.1 *(17 points)*

In this exercise we will use Linear Regression to estimate a model for our data.

a) Download the dataset Auto MPG from here:

https://archive.ics.uci.edu/ml/datasets/Auto+MPG

Load the first and third columns (mpg and displacement) which represent the Miles per galon and Engine displacement in cubic inches of various car models. They will act as our $\mathbf{X}$ and $\mathbf{Y}$ respectively for this exercise. Use the first **50** point pairs as a training set, and the next **50** will be the test set. *(1 point)*

b) Plot the training data points. What is your estimation for the degree of the polynomial that would approximate this data? *(1 point)*

c) We will fit a first order model to this training set. State the equation for such a model and the gradient descent update rule. *(2 points)*

d) Implement gradient descent to find the parameters of the model. Choose a suitable learning rate and initialization. Plot the value of the cost function per epoch. *(4 points)*

e) Using the parameters that your algorithm converged to, plot the model over the data. What is final value of the cost function? *(1 point)*

f) Now we will fit a second order model. Again, state the equations for the model and the gradient descent. *(1 point)*

g) Find the parameters of that model using gradient descent. Plot your model over the data and also plot the value of the cost function per epoch. What is the final value of the cost function in this case? Is it expected? *(3 points)*

h) Use numpy's polyfit in Python, or a similar function, to fit a 9th degree polynomial. We will use that polynomial to evaluate the performance of our model (You can use numpy's polyval). *(1 point)*

i) Use the 3 models you estimated (first, second, and 9th order polynomials) to evaluate the prediction values $\mathbf{Y}$ for the test set. Calculate the mean square error (MSE as defined in the lecture) for each case. What do you observe? *(3 points)*

Regularization

Exercise 3.2 *(3 points)*

In the lecture (slide 15) We saw the equation.

$$\mathrm{J}(\mathbf{w}) = \mathrm{MSE}_{\mathrm{train}} + \lambda\, \mathrm{w}^{\top}\, \mathrm{w}$$

- Drive a closed form expression for the regularization term given **w** from slide 9 in the lecture.

Submission instructions

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

Submission architecture

You have to generate a **single ZIP file** respecting the following architecture:

```
tutorial2_<matriculation_nb1>_<matriculation_nb2>_<matriculation_nb3>
|
+--- source
|      |
|      +------- file 1
|      +------- file 2
|      +------- ...
+--- report.pdf
+--- README.txt
```

where

- **source** contains the source code of your project,
- **report.pdf** is the report where you present your solution with **the explanations** and the plots.
- **README** which contains group member informations (name, matriculation numbers and emails) and a **clear** explanation about how to compile and run your source code

The ZIP filename has to be :

```
tutorial2_<matriculation_nb1>_<matriculation_nb2>_<matriculation_nb3>.zip
```

Some hints

We advice you to follow the following guidelines in order to avoid problems :

- Avoid building complex systems. The exercises are simple enough.
- Do not include any executables in your submission, as this will cause the e-mail server to reject it.

Grading

Send your assignment to the tutor who is responsible of your group:

- Merlin Köhler s9mnkoeh@stud.uni-saarland.de
- Yelluru Gopal Goutam goutamyg@lsv.uni-saarland.de
- Ahmad Taie s8ahtaie@stud.uni-saarland.de

The email subject should start with `[PSR TUTORIAL 3]`