Universität des Saarlandes
FR 4.7 Sprach- und Signalverarbeitung

Prof. Dietrich Klakow
Pattern and Speech Recognition

Exercise Sheet 6

Regularization for Deep Learning

Deadline: 03.01.2017, 23:59

Exercise 6.1 *(2 points)*

For each of the following statements state if it's true or false and justify your answer

a) Regularization is any modification we make to a learning algorithm that is intended to reduce its generalization error as well as its training error.

b) The model parameters α can be learned using linear regression for the model: $y = \log(x^{\alpha_1} e^{\alpha_2})$

Exercise 6.2 *(2 points)*

In what way is Dropout similar to Bagging and how do they defer? What about Boosting and Bagging? (Your answer should not exceed 4 sentences)

Exercise 6.3 *(5 points)*

Consider a quadratic error function of the form

$$E = E_0 + \frac{1}{2}(w - w^*)^T H (w - w^*)$$

where w^* represents the minimum, and the Hessian matrix H is positive definite and constant. Suppose the initial weight vector $w^{(0)}$ is chosen to be at the origin and is updated using simple gradient descent

$$w^\tau = w^{\tau-1} - \epsilon \nabla E$$

where τ denotes the step number, and ϵ is the learning rate (which is assumed to be small). After τ steps, the components of the weight vector parallel to the eigenvectors of H can be written

$$w_j^\tau = \{1 - (1 - \epsilon\lambda_j)^\tau\} w_j^*$$

where $w_j = w^T v_j$ and v_j and λ_j are the eigenvectors and eigenvalues, respectively, of H.

a) Show that as $\tau \to \infty$, this gives $w_\tau \to w^*$ as expected, provided $|1 - \epsilon\lambda_j| < 1$.

b) Now suppose that training is halted after a finite number of τ steps. Show that the components of the weight vector parallel to the eigenvectors of the Hessian satisfy

$$w_j^\tau \simeq w_j^* \text{ when } \lambda_j \gg (\epsilon\tau)^{-1}$$

$$|w_j^\tau| \ll |w_j^*| \text{ when } \lambda_j \ll (\epsilon\tau)^{-1}$$

c) Show that $(\epsilon\tau)^{-1}$ is analogous to the regularization parameter α from slide 6 in L_2 Parameter Regularization.

Exercise 6.4 *(11 points)*

a) Download the dataset from the course website. Plot the data, where the first and second columns are the 2d features (u, v), and the third column is the class. Give each class a different representation. *(1 points)*

b) We want to fit a regularized logistic regression model to the data. Recall the hypothesis function is

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

Start by coding a function that builds the feature vector x from all monomials of u and v up to the sixth power. *(1 points)*

c) The cost function for the logistic regression with L_2 regularization is

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)})) + (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

Where λ is the regularization parameter and y is the class. Derive the gradient and the hessian for this cost function. (You don't have to write the full vector/matrix) *(2 points)*

d) Implement Newton's method using the equations you derived in the previous question. *(4 points)*

e) Plot the decision boundary (Evaluate $\theta^T x$ over a grid of points (u, v) and plot $\theta^T x = 0$) for λ with values 0, 1, and 10. Comment on the plots. *(3 points)*

Christmas *Bonus*

Exercise 6.5 *(1 points)*

Santa wants to use a Single-layer perceptron to decide who has been naughty and who has been nice, Which of the following can he do with it?

a) A person is nice if they helped an old person or gave food to a beggar.

b) A person is naughty only if they pick their nose and they don't finish their food.

c) A person is nice if they like Santa or the Grinch, but not both.

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 ataie@lsv.uni-saarland.de

The email subject should start with `[PSR TUTORIAL 6]`