Introduction to ML
Assignment no. 2

Due: April 15th, 2018 (no extensions)

1 Submission Instructions:

1. **Theoretical Part**

   (a) Write your full name and ID at the top of your solution.

   (b) Submission: please submit a PDF version named `Ex_2_theoretical.pdf` via the Submit system.

2. **Practical Part**

   (a) You are not allowed to use any machine learning packages or tools (e.g. scikit-learn, PyBrain, PyML, etc.).

   (b) You are allowed to use numpy package

   (c) Use Python 2.7

   (d) In order to submit your solution - please submit the relevant files to the corresponding assignment on the Submit system.

   Your files should include:

   i. A text file called `details.txt` with your full name (in the first line) and ID (in the second line).

   ii. `ex2.py` - Your code file.

Good Luck!

2 Theoretical Part:

1. Multiclass and logistic regression
   Let $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_m, y_m)\}$ be a training set of examples, where $\mathbf{x}_i \in \mathbb{R}^d$ is an input feature vector of length d and $y_i \in \{1, \cdots, k\}$ is the output, where k is the number of classes.

   (a) Define the logistic regression conditional probability $P(Y = y|X = \mathbf{x})$ for multiclass classification.

   (b) Express the optimization problem that minimizes the negative log likelihood over the training set.

   (c) Find the update rule of this optimization problem using stochastic gradient descent (SGD) (*Hint*: find the gradients for all the relevant parameters, for all classes).

3 Practical Part:

1. Assume input x is scalar and there are 3 classes with equal priors. Each conditional density is normal:

$$f(x \mid y = a) = \mathcal{N}(2a, 1), \qquad a = 1, 2, 3. \tag{1}$$

   Sample 100 points from each class and train a logistic regression based on this training data. Plot on the same graph the estimated posterior probability $p(x|y = 1)$ based on the logistic regression you trained and the posterior probability based on the true distribution:

$$p(x \mid y = 1) = \frac{f(x \mid y = 1)}{f(x \mid y = 1) + f(x \mid y = 2) + f(x \mid y = 3)} \tag{2}$$

   Draw the graph for x in the range $[0, 10]$ and add it to `Ex_2_theoretical.pdf`.