50.007 Machine Learning, Fall 2020

Homework 2: SVM and Logistic Regression

Due: 8 November 2020

This homework will be graded by Perry Lam.

1 Support Vector Machines

Question 1.1 [15 pts]

Given the mapping $x = [x_1 \ \ x_2]^T \rightarrow \varphi(x) = [1 \ \ x_1^2 \ \ \sqrt{2}x_1x_2 \ \ x_2^2 \ \ \sqrt{2}x_1 \ \ \sqrt{2}x_2]^T$

a) Determine the Kernel $K(x, y)$ [*10pts*]

b) Calculate the value of the Kernel $K(x, y)$ if $x = [1 \ \ 2]^T$ and $y = [3 \ \ 4]^T$ [*5pts*]

Question 1.2 [20 pts]

The primal problem of SVM with soft margin is given below:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}w^T w + C\sum_{i=1}^{N} \xi_i \\ &\text{subject to} \quad d_i(w^T x_i + b) - 1 + \xi_i \geq 0 \ , \quad \xi_i \geq 0 \end{aligned}$$

1) Using Lagrange multipliers and KKT conditions, can you derive the formulation of dual problem with soft margin? Please note that the dual form is already provided in slides, so we expect you to go through the mathematical steps. [*15pts*]

2) Explain in which cases we would prefer to use soft margin rather than hard margin. [*5pts*]

Question 1.3: Hands-on [15 pts]

Download and install the widely used SVM implementation LIBSVM `https://github.com/cjlin1/libsvm`, or `https://www.csie.ntu.edu.tw/~cjlin/libsvm/`. We expect you to install the package on your own – this is part of learning how to use off-the-shelf machine learning software. Read the documentation to understand how to use it.

Download sonar folder. In that folder there are training.txt and test.txt, which respectively contain 145 training examples and 63 test examples in LIBSVM format. The task is to train an SVM model to discriminate between sonar signals, and to classify if the object is a rock or a mine (metal cylinder) given 60 attributes about the signal. This is a binary classification task.

Run LIBSVM to classify objects with the following kernels: a) linear, b) polynomial, c) radial basis function, and d) sigmoid. You should use default values for all other parameters. Please report the test accuracy for each kernel. Which kernel would you choose and why?

2 Logistic Regression

Question 2.1 [20 pts]
Suppose that you have trained a logistic regression classifier, and it outputs on a new example a prediction $h_\theta(x) = 0.35$. This means (check all that apply):
1) Our estimate for $P(y = 0|x; \theta)$ is 0.35
2) Our estimate for $P(y = 0|x; \theta)$ is 0.65
3) Our estimate for $P(y = 1|x; \theta)$ is 0.35
4) Our estimate for $P(y = 1|x; \theta)$ is 0.65
Please explain your answer for each point.

Question 2.2 [10 pts]
Suppose you train a logistic classifier $h_\theta(x) = g(\theta_0 + x_1\theta_1 + x_2\theta_2)$, and obtain $\theta = [6 \;\, -1 \;\, 0]^T$. Please formulate the decision boundary of your classifier. Note that this is a binary classification problem, which means class label y can be 0 or 1.

Question 2.3 [10 pts]
Suppose you train a logistic classifier $h_\theta(x) = g(\theta_0 + x_1\theta_1 + x_2\theta_2 + x_1^2\theta_3 + x_2^2\theta_4)$, and obtain $\theta = [-9 \;\, 0 \;\, 0 \;\, 1 \;\, 1]^T$. Please formulate the decision boundary of your classifier. Note that this is a binary classification problem, which means class label y can be 0 or 1.

Question 2.4 [10 pts]
In logistic regression, we find the parameters of a logistic (sigmoid) function that maximize the likelihood of a set of training examples. The likelihood is given as follows:

$$\prod_{i=1}^{n} P(y^{(i)}|x^{(i)}) \tag{1}$$

However, we re-express the problem of maximizing the likelihood as minimizing the following expression:

$$\frac{1}{n}\sum_{i=1}^{n} log\left(1 + exp\left(-y^{(i)}\left(\theta.x^{(i)} + \theta_0\right)\right)\right) \tag{2}$$

What is the benefit of optimizing the log-likelihood rather than the likelihood of the data? In other words, why is this expression computationally more "convenient"? *(Hint: try randomly generating, say, 1,000 probabilities in Python and multiplying them together as in Eq. 1.)*