Machine Learning 2009
(Home work 1)

May 18, 2009

- Due on Monday, June 1, before 10 a.m.
- Late submissions will not be accepted.
- Submit hard copy of the results, plots and your workings
- Submit a printed copy of the codes also.
- You may save time if you use MATLAB for the computations and plots.
- Please do not hesitate to contact me if you do not understand the problems.

1. **[20 points + 20 points + 10 points + 10 points] Regression and Logistic Regression**

   (a) Given the real-valued function $f(x_1, x_2, ..., x_n)$, if all partial second derivatives of f exist, then the Hessian matrix of f is the matrix

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

   i. Let $J(\boldsymbol{\theta})$ be the usual cost function for linear regression for a data in $\Re$, i.e.,

$$J(\theta_0, \theta_1) = \sum_{i=1}^{n} ((\theta_0 + \theta_1 x^{(i)}) - y^{(i)})^2.$$

   Compute the Hessian matrix H for $J(\boldsymbol{\theta})$. Show that H is positive semi-definite, i.e., for any $\boldsymbol{z} \in \Re^2$, $\boldsymbol{z}^T H \boldsymbol{z} \geq 0$.

ii. Consider logistic regression for a data set

$$X = \{(\boldsymbol{x}^{(i)}, y^{(i)}) : \boldsymbol{x}^{(i)} \in \Re^2, y^{(i)} \in \{0, 1\}, i = 1, \ldots, m\}.$$

The hypothesis we use in logistic regression is

$$h_\theta(\boldsymbol{x}) = \frac{1}{1 + \exp[-(\theta_0 + \theta_1 x_1 + \theta_2 x_2)]}.$$

Thus, the likelihood function can be defined as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{m} \left[y^{(i)} \log(h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})) \right].$$

Calculate the Hessian matrix for this function and show that it is negative semi-definite.

Note: These shows that that the cost function for linear regression and likelihood function for logistic regression are convex and concave respectively. It implies that these functions do not have any local optimas. The cost function for regression has global minima and that of logistic regression has a global maxima. This result is true for higher dimensions also You may try proving it, but no points for it.

(b) Consider logistic regression for a data set

$$X = \{(\boldsymbol{x}^{(i)}, y^{(i)}) : \boldsymbol{x}^{(i)} \in \Re^n, y^{(i)} \in \{0, 1\}, i = 1, \ldots, m\}.$$

The hypothesis we use in logistic regression is

$$h_\theta(\boldsymbol{x}) = \frac{1}{1 + \exp[-(\theta_0 + \sum_{i=1}^{m} \theta_i x_i)]}.$$

Now we define a modified likelihood function as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{m} \left[y^{(i)} \log(h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})) + (1 - y^{(i)}) \log(1 - h_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)})) \right] - \frac{\lambda}{2} \sum_{i=0}^{n} \theta_i^2.$$

Note that this likelihood function differs from the usual one by the extra term $(1/2)\lambda \sum_{i=0}^{n} \theta_i^2$. This version of the likelihood function is called the L_2 regularized likelihood. The maximization of this function penalizes those values of $\boldsymbol{\theta}$ which are large. The parameter λ is a constant called the regularization parameter. Derive the online gradient descent rule for this likelihood function.

(c) Download the data data.dat from the course website. Use the online gradient descent rule to fit a line on the data. Plot the points along with the line. Give the equation of the fitted line.

(d) Download the data class.txt from the course website. Plot the points, you should use a different marker for denoting points in the two different classes. Now, plot the decision boundary obtained by logistic regression. Give the equation of the decision boundary.

2. **[10 +10 +10 +10 points] Bayesian Decision Theory**

(a) Let x have an exponential density:

$$p(x|\theta) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Suppose n samples $x_1, x_2, \ldots, x_n$ are drawn independently according to $p(x|\theta)$. Show that the maximum likelihood estimate for θ is given by

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i}$$

(b) Consider a two class classification problem with two classes c_1 and c_2 with the following prior and class conditional distributions: $P(c_1) = P(c_2) = 0.5$, $p(\boldsymbol{x}|c_1) = \mathcal{N}(\boldsymbol{\mu}_1, \Sigma_1)$ and $p(\boldsymbol{x}|c_2) = \mathcal{N}(\boldsymbol{\mu}_2, \Sigma_2)$. Derive the equation of the Bayesian discriminant functions for the following values of μ and Σ. Plot the decision boundaries.

- $\boldsymbol{\mu}_1 = [2, 8]^T$, $\boldsymbol{\mu}_2 = [8, 2]^T$, $\Sigma_1 = \Sigma_2 = \begin{bmatrix} 3.0 & 0.0 \\ 0.0 & 3.0 \end{bmatrix}$
- $\boldsymbol{\mu}_1 = [3, 6]^T$, $\boldsymbol{\mu}_2 = [3, -3]^T$, $\Sigma_1 = \begin{bmatrix} 0.5 & 0.0 \\ 0.0 & 2.5 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2.0 & 0.0 \\ 0.0 & 2.0 \end{bmatrix}$
- $\boldsymbol{\mu}_1 = [3, 6]^T$, $\boldsymbol{\mu}_2 = [3, -3]^T$, $\Sigma_1 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 2.5 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2.0 & 0.5 \\ 0.5 & 2.0 \end{bmatrix}$

(c) Download the file class2.txt from the course website. It has data for a two class classification problem. The features are in $\Re^2$ and the classes are denoted by 0 and 1. Assume that the data from the two classes are generated from a normal distribution. Estimate the prior probabilities and the class conditional densities for the two classes.

(d) Build a Bayes Classifier using the probability values obtained from the previous problem. Now download the file class3.txt. Ignore the labels in the data and classify the data points using the classifier. Use the class labels to compute the number of misclassifications done by your classifier.