Homework 1

MACHINE LEARNING — 15 October, 2013
1st Semester (FEUP-PDEEC) — 18-782PP CMU-Portugal

Submit by 29 October, 2013, 23h59
by email to `jaime.cardoso@fe.up.pt`
and `vikramg@andrew.cmu.edu`

Problem 1

Consider a binary classification problem, where we want to label $y_i \in \{0, 1\}$ to observation x_i. Let $\rho_i = p(Y_i = 1|x_i)$ and $1 - \rho_i = p(Y_i = 0|x_i)$. Suppose that the classifier is allowed to make one of three decisions: choose class 0, choose class 1, or *reject* this data (refuse to make a decision). We can use bayesian approach to tradeoff the losses incurred by incorrect decisions and rejections. Suppose that the classifier incurs a loss of 0 when it chooses the correct class, a loss of 1 when it chooses a wrong class, and a loss of λ whenever it selects the reject option. Express the optimal decision rule, which minimizes the posterior expected loss, as a function of ρ_i and $\lambda \geq 0$.

Problem 2

1. Suppose we have a very large training set (say $N = 1000000$) but only two classes ($r = 2$), only a few features (say $d = 4$) and only a few values per feature (say $q = 5$). Explain how we can train a Bayesian classifier without making the *naive* Bayes assumption. (Hint: How many parameters do we need to learn, if we dont make the naive Bayes assumption?)

2. Given the following data about cellphones with various attributes such as gender of the owner, color and type and the corresponding information about whether they lasted for more than one year or not. Find out (using Bayes classifier) that whether a white colored phone with a touchscreen owned by a male (i.e. with attributes specified as: <Male, White, Touchscreen>) is more probable to last more than one year or not. Make suitable assumptions if needed, and specify them clearly in the solutions.

User Gender	Color of Phone	Type	Lasted one year?
Male	Black	Touchscreen	Yes
Male	Black	Touchscreen	No
Male	Black	Touchscreen	Yes
Female	Black	Touchscreen	No
Female	Black	Keypad	Yes
Female	White	Keypad	No
Female	White	Keypad	Yes
Female	White	Touchscreen	No
Male	White	Keypad	No
Male	Black	Keypad	Yes

Homework 1

MACHINE LEARNING — 15 October, 2013
1st Semester (FEUP-PDEEC) — 18-782PP CMU-Portugal

Problem 3

1. Implement in Matlab a function
   `[weights] = regression(train_data_X, train_data_Y)`
   that outputs the coefficients of the linear regression on the data. Test it to compute the polynomial regression of degree 9 with the data set Sin(x) from PRMLBishop. Obtain figure 1.4.d from PRMLBishop.

2. In some contexts it is interesting to introduce different costs per example in the error function
   $$\mathcal{L}(\mathbf{w}, S) = \frac{1}{2}\sum_{n=1}^{N} c_n(y_n - \mathbf{w}^t\mathbf{x}_n)^2 + \frac{1}{2}\lambda\mathbf{w}^t\mathbf{w}$$
   Deduce in matrix notation the normal equation.

3. Adapt the function written in a) to output the coefficients as given by b). The function should have two new input parameters: a vector c and a scalar λ.