Machine learning: Minor 1

Instructor - Jayadeva / Prathosh

9th November 2020

Total Marks: 20 Time: 60 Minutes

1. Suppose you are tasked to a design a email spam filter where every email has equal chance of being a spam email or otherwise. Suppose there are d words in a dictionary that are deemed as spam words. Let the occurrence of each spam word in an email is independent of that of every other spam word. Also assume that probability of occurrence of any spam word in a spam email is p $>$ 0.5 and the probability of it occurring in a non-spam email is 1$-$p. Assuming d to be odd, derive the rule for the optimal spam filter. Hint: A Naive Bayes classifer is a Bayes classifier in which the features of the data (X) are statistically independent.

2. Solve the following questions:

   (a) Suppose we have a binary classification problem with $X \in S \subset R^2, S = X : |X_1|, |X_2| \leq 2$ where X_1 and X_2 are feature dimensions. Let $y \in 0, 1$ and true classes are y $=$ 1 if X1 $>$ 0 and y $=$ 0 otherwise. Suppose there is a classifier $h(X) = 1$ if $X_1^2 + X_2^2 > 1$ and $h(X) = 0$ otherwise. Calculate the probability of the misclassification of this classifier.

   (b) Consider a binary classification problem with scalar features. Given the following with the usual notations, write an expression for the optimal classifier in terms of the features -L (0, 0) $=$ L (1, 1) $=$ 0, L (1, 0) $=$ 2 , L (0, 1) $=$ 4, $p_0 = 0.4$ and $p_1 = 0.6$, $f_0(x) \sim \mathcal{N}(\mu_0, \Sigma)$ and $f_1(x) \sim \mathcal{N}(\mu 1, \Sigma)$.

   (c) State the difference between the ML, MAP estimates.

   (d) Suppose we have drawn n IID samples from a uniform distribution between 0 and $a \in Z^+$, derive the ML estimate.

3. Let $L(i, j)$ denote the loss of a classifier when it decides class $i$and the true class is j. Suppose there are K classes and the $K + 1$ st class corresponds to rejecting the pattern. The loss function is defined as

$$
\begin{aligned}
L(i, j) = 0 \quad & if \quad i = j \quad and \quad i, j = 1, ..., K \\
= \rho_m \quad & if \quad i = 1, ..., K, \quad and \quad i \neq j \\
= \rho_r \quad & if \quad i = K + 1
\end{aligned}
$$

Show that the optimal classifier in this case is that which decides as follows - Choose class i if $[q_i(X) \geq q_j(X), \forall j \neq i) \quad AND \quad (q_i(X) \geq 1 - (\rho_r/\rho_m)]$; otherwise reject (class $K+1$), where $q_i(X)$ is the posterior of class i at input X. Also explain what this decision means when $\rho_r = 0$, $\rho_r > \rho_m$.

4. Define a conjugate prior. Derive the MAP estimate for the success probability parameter of a Bernoulli RV. Hint: Beta density with parameters a, b is given by $f(p) = [\Gamma(a+b)/\Gamma(a)\Gamma(b)]p^{a-1}(1-p)^{b-1}$; $p \in [0,1]$, a, b ≥ 1 , $\Gamma()$ denoting the Gamma function. Mode of the beta density is at $(a-1)/(a+b-2)$. Compare the MAP estimate with corresponding ML estimate and contrast.

5. Solve the following:

   (a) Suppose the cost function of theLinear Regression is modified as following:

   $$||Y - \Phi(X)W||_2 + \alpha||W||_2^2$$

   Derive the optimal estimate for W in this case and relate it to ordinary least squares.

   (b) Derive the optimal estimate for the intercept term in the linear regression and interpret it.

   (c) How do you extend a linear discriminate function to multi-class classification problem.