Assignment 3: Linear Classifcation

CS489/698 – Winter 2010

Out: March 2, 2010
Due: March 16, 2010

Be sure to include your name and student number with your assignment.

1. **[32 pts]** Linear separability

   (a) **[16 pts]** Given a set of data points $\{u_i\}$, we can define the convex hull to be the set of all points x given by

   $$u = \sum_i \alpha_i u_i$$

   where $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. Consider a second set of points $\{v_j\}$ together with their corresponding convex hull. By definition, the two sets of points will be linearly separable if there exists a vector w and scalar b such that $w^T u_i + b > 0$ for all u_i and $w^T v_j + b < 0$ for all v_j. Show that if their convex hulls intersect, the two sets of points cannot be linearly separable, and conversely that if they are not linearly separable, their convex hulls intersect.

   (b) **[16 pts]** Consider a threshold perceptron that predicts $y = 1$ when $w^T x + b \geq 0$ and $y = 0$ when $w^T x + b < 0$. It is interesting to study the class of Boolean functions that can be represented by a threshold perceptron. Assume that the input space is $X = \{0, 1\}^2$ and the output space is $Y = \{0, 1\}$. For each of the following Boolean functions, indicate whether it is possible to encode the function as a threshold perceptron. If it is possible, indicate some values for w and b. If it is not possible, indicate a feature mapping $\phi : X \rightarrow \hat{X}$ with values for w and b such that $w^T \phi(x) + b$ is a linear separator that encodes the function.

   - and
   - or
   - exclusive-or
   - iff

2. **[18 pts]** Prove the following properties of the logistic sigmoid function σ:

   - $\sigma(-a) = 1 - \sigma(a)$
   - $\sigma^{-1}(a) = ln(a/(1-a))$
   - $\frac{\partial \sigma}{\partial a} = \sigma(a)(1 - \sigma(a))$

3. **[50 pts]** Linear models for classication. Implement the following two classification algorithms. A dataset will be posted on the course web page. The input space is continuous (i.e., $X = \Re^d$), while the output space is categorical (i.e., $Y = \{C_1, C_2\}$). Learn the parameters of each classifier by likelihood maximization.

   (a) **[25 pts]** Mixture of Gaussians: let $\pi = \Pr(y = C_1)$ and $1 - \pi = \Pr(y = C_2)$. Let $\Pr(x|C_1) = N(x|\mu_1, \Sigma)$ and $\Pr(x|C_2) = N(x|\mu_2, \Sigma)$. Learn the parameters π, μ_1, μ_2 and Σ by likelihood mazimization with the training data. Use Bayes theorem to compute the probability of each class given an input x: $\Pr(C_j|x) = k \Pr(C_j) \Pr(x|C_j)$.

   (b) **[25 pts]** Logistic regression: let $\Pr(C_j|x) = \sigma(w_j^T x + b_j)$. Learn the parameters w and b by likelihood maximization with the training data. More specifically use Newton's algorithm derived in class to optimize the parameters.