Assignment 2: Mixtures of Gaussians and Logistic Regression

CS489/698 – Winter 2018

Out: January 22, 2017
Due: February 2 (11:59pm)

Submit an electronic copy of your assignment via LEARN. Late submissions incur a 2% penalty for every rounded up hour past the deadline. For example, an assignment submitted 5 hours and 15 min late will receive a penalty of ceiling(5.25) * 2% = 12%.

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

1. **[50 pts]** Implement the following two classification algorithms. Use the same dataset (handwritten digits) as for assignment 1 to train the algorithms. Test the algorithms by 10-fold cross validation.

   (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 maximization. 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 $\Pr(C_1|x) = \sigma(\mathbf{w}^T\mathbf{x} + w_0)$ and $\Pr(C_2|\mathbf{x}) = 1 - \sigma(\mathbf{w}^T\mathbf{x} + w_0)$. Learn the parameters $\mathbf{w}$ and w_0 by conditional likelihood maximization. More specifically use Newton's algorithm derived in class to optimize the parameters. 10 iterations of Newton's algorithm should be sufficient for convergence. You do not need to regularize logistic regression for this question.

   What to hand in:

   - Report the accuracy of mixtures of Gaussians and logistic regression obtained by 10-fold cross validation. Measure the accuracy by counting the average number of correctly labeled images. An image is correctly labeled when the probability of the correct label is greater than 0.5.
   - Briefly discuss the results:
     - Mixture of Gaussians and logistic regression both find a linear separator, but they use different parameterizations and different objectives. What can you conclude about the parameterizations and the objectives?
     - Mixture of Gaussians and logistic regression find a linear separator where as k-Nearest Neighbours (in assignment 1) finds a non-linear separator. What can you conclude about the different separators?
   - Print the parameters found for each model. Note that each run of cross validation will give you different parameters. Do not average them. Instead report the parameters found when re-training on *all* the data. While the accuracy reported by 10-fold cross validation is not for those parameters, the accuracy should be at least as good when you train on more data (i.e., all the data).
   - Print your code.

   (see next page for Question 2)

2. **[50 pts]** Linear separability

   (a) **[25 pts]** Consider a threshold perceptron that predicts $y = 1$ when $\mathbf{w}^T\mathbf{x} + w_0 \geq 0$ and $y = 0$ when $\mathbf{w}^T\mathbf{x} + w_0 < 0$. It is interesting to study the class of Boolean functions that can be represented by a threshold perceptron. Assume that the input space is $\mathbf{X} = \{0, 1\}^2$ 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 $\mathbf{w}$ and w_0. If it is not possible, indicate a feature mapping $\phi : X \rightarrow \hat{X}$ with values for $\mathbf{w}$ and w_0 such that $\mathbf{w}^T\phi(\mathbf{x}) + w_0$ is a linear separator that encodes the function.

   - and
   - or
   - exclusive-or
   - iff

   (b) **[25 pts]** Is the dataset used in Question 1 linearly separable? To answer this question, design an experiment that involves a logistic regression classifier that will allow you to verify whether the dataset is linearly separable. Describe your experiment and the results. What can you conclude from the results?