Occidental College

Department of Computer Science

CS 347 - Machine Learning

Assignment 2

You may work in group of 4 at most. Initial Due date: Sep 29 | Final Due date: Oct 2

Problem set

1. In this problem you will implement a special version of logistic regression. This logistic regression problem is to maximize:

$$J(\theta) = -\frac{\lambda}{2}\|\theta\|^2 \sum_{i=1}^{m} w^{(i)}[y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))]. \tag{1}$$

Here, $w^{(i)}$ are pre-set weights. Some relevant equations: The gradient of $l(\theta)$ is given by:

$$\nabla_\theta J(\theta) = X^T w^i [y^{(i)} - h_\theta(x^{(i)})] - \lambda\theta \tag{2}$$

The Hessian is given by:

$$H = X^T D X - \lambda I \tag{3}$$

where I is the identity matrix, and D is a diagonal matrix with:

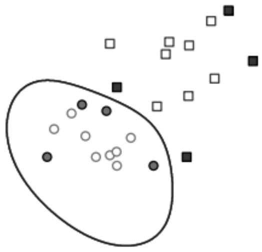
$$D_{ii} = -w^{(i)} h_\theta(x^{(i)})(1 - h_\theta(x^i)) \tag{4}$$

(they are not so difficult to derive) Given a point x_q. Suppose we choose weight:

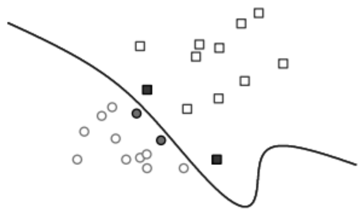
$$w^{(i)} = \exp(\|x_q - x^{(i)}\|^2/(2\tau^2)) \tag{5}$$

- Use the dataset provided, implement $y_{pred} = RegressionAtHome(X_{train}, Y_{train}, x_{pred}, tau)$, It should do the following: Compute the weight $w^(i)$ for each training example, use Newton's method to optimize $J(\theta)$ and return an output $y = 1$ if $h_\theta(x) > 0.5$).
- Plot decision boundary by doing the following: Create a grid $n \times n$ of the graph. For each cell, run $RegressionAtHome$. Start with 50×50 and gradually raise the resolution until it takes too long (longer than 15 miuntes is probably too long). Since it is a probabilistic model, you should use heat map to demonstrate the decision boundary.
- Plot the decision boundary for each of parameter $\tau = 0.01, 0.04, 0.1, 0.5, 1$. What happen as τ increase?
- Apply the scheme to the iris dataset. You just need to classify 2 classes: 0 and 1. Choose 2 appropriate values for τ. Each τ should have 4 decision boundary plot that are combinations of petal length and width, and sepal length and width.
- (optional) Explain what the scheme try to achieve.

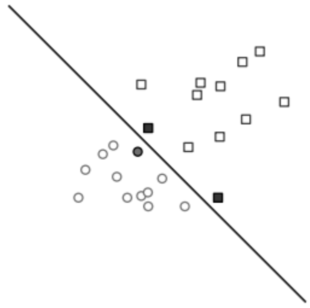
2. One thing I forgot to mention in lecture, support vectors are the data points that are closest to the separating hyperplane and essentially "support" or define the hyperplane. In hard-margin SVM, these are the data points that has the geometric margin distance away from the decision boundary. In soft-margin SVM, these are data points that lies within the geometric margin.



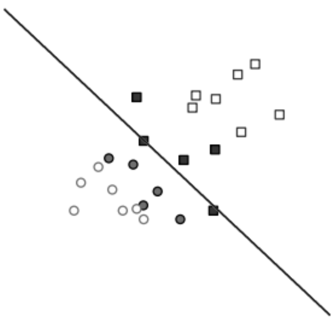
1



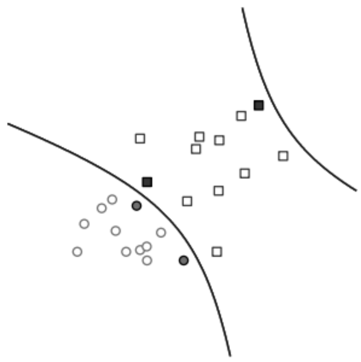
2



3



4



5

6

Above, there are different SVMs with different shapes/patterns of decision boundaries. The training data is labeled as $y_i \in \{-1, 1\}$, represented as the shape of circles and squares respectively. Support vectors are drawn in solid circles. Match the scenarios described below to one of the 6 plots (note that one of the plots does not match to anything). Each scenario should be matched to a unique plot. Explain in less than two sentences why it is the case for each scenario.

- A soft-margin linear SVM C=.02
- A soft margin linear SVM C=20
- A hard-margin kernel SVM with $k(u, v) = uv + (uv)^2$
- A hard margin kernel SVM with $k(u, v) = \exp(-5\|u - v\|^2)$
- A hard margin kernel SVM with $k(u, v) = \exp(-\frac{1}{5}\|u - v\|^2)$