Problem Set 1

September 29, 2016

1. Naive Bayes

Solution:

1

$$\begin{aligned}\hat{y}_{x_{new}} &= \text{argmax}_k \hat{p}(y = k|x_{new}) \\ &= \text{argmax}_k \hat{p}(x_{new}|y) \cdot \hat{p}(y = k) \\ &= \text{argmax}_k \prod_{i=1}^{5} g(x_{new}^{(i)}|\hat{\mu}_{ki}, \hat{\sigma}_{ki})\, \hat{p}(y = k)\end{aligned}$$

2 Let J_k be those indices j such that $y_j = k$. Then $|J_k|$ is the number of examples of class k.

$$\hat{\mu}_{ki} = \frac{1}{|J_k|} \sum_{j \in J_k} x_j^{(i)}$$

$$\hat{\sigma}_{ki} = \frac{1}{|J_k|} \sum_{j \in J_k} (x_j^{(i)} - \hat{\mu}_{ki})^2$$

$$\hat{p}(Y = y) = \frac{|J_k|}{n}$$

3 The Naive Bayes assumption is that the dimensions of each data point x are conditionally independent given y. This is the same as taking the covariance matrix of each class conditional to be diagonal.

2. Linear Classifier

Consider a perceptron classifier in $\mathbb{R}^2$, given by a hyperplane with orthogonal vector v_H and offset c and two points $\mathbf{x}_1$ and $\mathbf{x}_2$. Suppose that

$$v_H := \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad c := \left(\frac{1}{2\sqrt{2}}\right) \mathbf{x}_1 = \begin{pmatrix} -3 \\ 0 \end{pmatrix} \quad \mathbf{x}_2 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

1 Compute the classification result for $\mathbf{x}_1$ and $\mathbf{x}_2$.

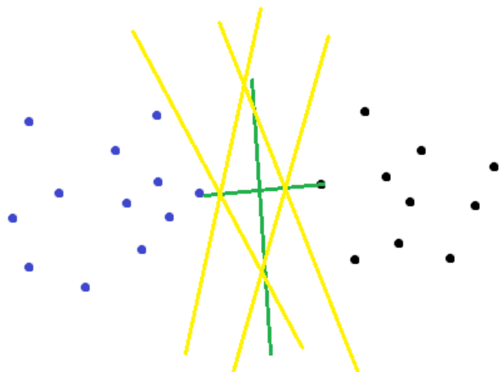
2 If a perceptron and SVM are trained on the same linearly separable classification problem, in what ways could the two classifiers disagree on the training data? On test data? Answer the second problem with an 2D illustration of trained classifiers, training data, and test data. Make sure both classifiers, as drawn, could actually be those learned by the perceptron and SVM, respectively.

Solution:

1 $class(x_1) = \text{sign}(\langle x_1, v_H \rangle - c) = -1$

$class(x_2) = \text{sign}(\langle x_2, v_H \rangle - c) = 1$

2 Perceptron and SVM will agree on the training data. They need not agree on the test data because they won't generally produce the same decision boundary.



Each of the yellow lines represents a decision boundary that could have been produced by perceptron. (There are many more... indeed, every decision boundary that separates the data could be the result of perceptron. To see this, note that if we start with a valid boundary, then perceptron will not modify it!) The green line is the maximum margin decision boundary produced by SVM.

Plot taken from Xu Huiyan.