Exam Pattern Recognition
Friday, January 31, 2019
17.00-20.00 hours

Question 1: True or False? (20 points)

1. T (as opposed to discriminative approaches, that only model the conditional distribution of the class label given the features)
2. F (no, just local minimum)
3. F (no, it is assumed that the covariance matrix is the same for each class)
4. F (points in direction of steepest *increase*)
5. T
6. F (No, in general they produce different decision boundaries. See lecture 50A).
7. T
8. F (e.g. there could be two identical feature vectors with different class labels in the training set)
9. T (see chapter 9 of the Bishop book).
10. T (this predicts the conditional mean of t for each value of x, and should therefore minimize the SSE)

Question 2: Logistic Regression (24 points)

(a) If the value of `ink` increases, and the value of `horbal` stays the same, then the probability that the digit is "1" decreases.

(b) This makes sense, because the digit "1" tends to consume less ink then the digit "0".

(c) For an all white image, we have `ink`=0, and `horbal`=0, so

$$p(t = 1 \mid \mathtt{ink} = 0, \mathtt{horbal} = 0) = \frac{e^{13.16}}{1 + e^{13.16}} \approx 1$$

(d) If $13.16 - 0.66 \times \texttt{ink} - 0.73 \times \texttt{horbal} \geq 0$ then predict 1, otherwise predict 0.

(e)

$$\begin{aligned} p(t=1) &= p(t^* \geq 0) \\ &= p(\mathbf{w}^\top \mathbf{x} + U \geq 0) \\ &= p(U \geq -\mathbf{w}^\top \mathbf{x}) \\ &= p(U \leq \mathbf{w}^\top \mathbf{x}) \end{aligned}$$

The last step is justified by the symmetry of the distribution of U around 0. It is crucial that you mention this justification.

(f) For notational convenience, define

$$F(u) = p(U \leq u),$$

and suppose that

$$F(u) = (1 + e^{-u})^{-1}$$

To obtain the density function corresponding to this cumulative density function, we take the derivative:

$$\begin{aligned} \frac{dF(u)}{du} &= -(1 + e^{-u})^{-2} \times e^{-u} \times -1 \\ &= \frac{e^{-u}}{(1 + e^{-u})^2} \end{aligned}$$

This is indeed the logistic distribution. Hence the logistic response function is the cumulative density function corresponding to the logistic distribution. Therefore

$$p(t=1) = p(U \leq \mathbf{w}^\top \mathbf{x}) = F(\mathbf{w}^\top \mathbf{x}) = (1 + e^{-\mathbf{w}^\top \mathbf{x}})^{-1}$$

Question 3: Support Vector Machines (20 points)

(a) Using the given formulas, we compute: $b = 1$, $w_1 = -2$, and $w_2 = 2$. So the decision boundary is:

$$1 - 2x_1 + 2x_2 = 0$$

(b) It is clear that the horizontal line $x_2 = 2.5$ has maximum margin. So the decision boundary (before scaling) is

$$-2.5 + 0x_1 + x_2 = 0$$

We now scale the coefficients so that $t_i y_i = 1$ for the points $\mathbf{x}_i$ closest to the decision boundary. The points now have $t_i y_i = 1.5$, so we have to divide the bias and weights by 1.5 to scale them correctly. This gives the decision boundary

$$-1\frac{2}{3} + \frac{2}{3}x_2 = 0$$

(c) The solution at (a) has $w_1 = -2$, $w_2 = 2$, and all points on the correct side of the margin boundary, so zero slack. This gives

$$E = \frac{1}{2}\mathbf{w}^\top\mathbf{w} + \sum_{n=1}^{N} \xi_n = 4 + 0 = 4$$

The solution at (b) has $w_1 = 0$, $w_2 = \frac{2}{3}$, and two points on the wrong side of the margin boundary, with slack values $1\frac{1}{3}$ each. This gives

$$E = \frac{1}{2}\mathbf{w}^\top\mathbf{w} + \sum_{n=1}^{N} \xi_n = \frac{2}{9} + 2\frac{2}{3} = 2\frac{8}{9}$$

Hence, the solution at (b) is preferred.

(d) Suppose the closest point has $t_i y_i > 1$. Then we can shrink b and $\mathbf{w}$ until $t_i y_i = 1$, still satisfying the constraints. This will reduce $E = \frac{1}{2}\mathbf{w}^\top\mathbf{w}$, and thus produce a better solution.