SUSTech

Machine-learning

Fall 2020

Homework 3

Liu Yuxi 2020 年 10 月 29 日

3.1. Show that maximization of the class separation criterion given by $m_2 - m_1 = w^T(m_2 - m_1)$ with respect ot w, using a Lagrange multiplier to enforce the constraint $w^T w = 1$, leads to the result that $w \propto (m_2 - m_1)$.

Solution. content...

3.2. Show that the Fisher criterion

$$j(w) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$$

can be written in the form

$$j(w) = \frac{w^w S_B w}{w^T S_W w}$$

Solution. content...

3.3. Consider a agenerative classification model for K classes defined by prior class probabilities $p(C_k) = \pi_k$ and general class-conditional dendities $p(\phi|C_k)$ where ϕ is the input feature vector. Suppose we are given a training data set$\{\phi_n, t_n\}$ where $n = 1, ..., N$, and t_n is a binary target vector of length K that uses the 1-of-K coding scheme, so that it has components $t_{nj} = I_{jk}$ if pattern n is from class C_k. Assuming that data points are drown independently from this model, show that the maximum-likelihood solution for the prior probabilities is given by

$$\pi_k = \frac{N_k}{N}$$

where N_k is the number of data points assigned to class C_k.

3.4. Verify the relation

$$\frac{d\sigma}{da} = \sigma(1 - \sigma)$$

for the derivative of the logistic sigmoid function defined by

$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

3.5. There are several possible ways in which to generalize the concept of linear discriminant functions from two classes to c classes. One possibility would be to use $(c-1)$ linear discrimminant functions, such that $y_k(x) > 0$ for inputs x in class C_k and $y_k(x) < 0$ for not in class C_k. By drawing a simple example in two dimensions for $c = 3$, show that this approach can lead to regions of x-space for which the classification is ambiguous. Another approach would be to use one discriminant functino $y_{jk}(x)$ for each possible pair of classes C_j and C_k, such that $y_{jk}(x) > 0$ for patterns in class C_j and $y_{jk}(x) < 0$ for patterns in class C_k. For c classes, we would need $c(c-1)/2$ discriminant functions. Again, by drawing a specific example in two dimensions for $c = 3$, show that this approach can also lead to ambiguous regions.

3.6. Given a set of data points $\{x^n\}$ we can define the convex hull to be the set of points x given by

$$x = \sum_n \alpha_n x^n$$

where $\alpha_n >= 0$ and $\sum_n \alpha_n = 1$. Consider a second set of point $\{z^m\}$and its corresponding convex hull. The two sets of points will be linearly separable if there exists a vector $\hat{w}$ and a scalar ω_0 such that $\hat{w}^T x^n + \omega_0 > 0$ for all x^n, and $\hat{w}^T z^m + \omega_0 < 0$ for all z^m. Show that, if their convex hulls intersect, the two sets of points cannot be linearly separable, and conversely that, if they are linearly separable, their convex hulls do not intersect.