Machine Learning

Homework 4

Not collected, not graded.

1 Linear separability

Given a set of data points $\{\mathbf{x}_n\}$, $\mathbf{x}_n \in \mathbb{R}^D$, we can define the *convex hull* to be the set

$$\mathrm{Conv}(\{\mathbf{x}_n\}) := \left\{ \sum_n \alpha_n \mathbf{x}_n \;\middle|\; \forall n\colon \alpha_n \geq 0 \wedge \sum_n \alpha_n = 1 \right\}.$$

Consider a second set of points $\{\mathbf{y}_m\}$, $\mathbf{y}_m \in \mathbb{R}^D$, and its respective convex hull, $\mathrm{Conv}(\{\mathbf{y}_n\})$. By definition, the two sets of points are linearly separable if there exists a vector $\mathbf{w}$ and a scalar w_0 such that

$$\forall \mathbf{x}_n\colon \mathbf{w}^T\mathbf{x}_n + w_0 > 0 \qquad \text{and} \qquad \forall \mathbf{y}_n\colon \mathbf{w}^T\mathbf{y}_n + w_0 < 0.$$

1. Show that if their convex hulls intersect, the two sets of points cannot be linearly separable
2. Conversely, show that if they are linearly separable, their convex hulls do not intersect.

2 Probabilistic generative models (for classification)

We first consider the two-class case. As seen in class, the posterior probability for class C_1 can be written as

$$p(C_1 \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid C_1)p(C_1)}{p(\mathbf{x} \mid C_1)p(C_1) + p(\mathbf{x} \mid C_2)p(C_2)}.$$

Let the class-conditional distribution be Gaussian:

$$p(\mathbf{x} \mid C_k) := \mathcal{N}(\mathbf{x} \mid \mu_k, \Sigma_k).$$

1. For shared covariance matrices $\Sigma_k = \Sigma$, it was said that the posterior probability is given by

   $$p(C_1 \mid \mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x} + w_0),$$

   where $\sigma(a) := 1/(1+e^{-a})$ is the sigmoid, and where

   $$\mathbf{w} = \Sigma^{-1}(\mu_1 - \mu_2), \qquad w_0 = -\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2 + \ln\frac{p(C_1)}{p(C_2)}.$$

   Derive that result yourself.
2. What happens if the covariance matrices are not shared, i.e. $\Sigma_1 \neq \Sigma_2$?
3. Show that for shared covariance matrices the decision boundaries are linear in input space.
4. What is the influence of the class priors $p(C_1)$ and $p(C_2)$ on the decision boundaries?
5. Show that the logistic sigmoid function $\sigma(a)$ satisfies:
   (a) $\sigma(-a) = 1 - \sigma(a)$,
   (b) $\sigma^{-1}(y) = \ln(y) - \ln(1-y)$.