

Statistical Methods for Artificial Intelligence, Autumn 2006
Problem set 5, Due Wednesday Nov. 15

Problem 1. This problem is on quadratic discriminant analysis. Consider a sample of pairs $\langle x_1, y_1 \rangle, \dots, \langle x_T, y_T \rangle$ with $\Phi(x_t) \in R^2$ and $y_t \in \{-1, 1\}$. Let $\Phi(x)$ be the pair of numbers $\langle f_1(x), f_2(x) \rangle$. Let S^+ and S^- be defined as follows.

$$S^+ = \{t : 1 \leq t \leq T \wedge y_t = 1\}$$

$$S^- = \{t : 1 \leq t \leq T \wedge y_t = -1\}$$

Now consider the following quantities.

$$\mu_1^+ = \frac{1}{|S^+|} \sum_{t \in S^+} f_1(x_t)$$

$$\mu_2^+ = \frac{1}{|S^+|} \sum_{t \in S^+} f_2(x_t)$$

$$\Sigma_{i,j}^+ = \frac{1}{|S^+|} \sum_{t \in S^+} (f_i(x_t) - \mu_i^+)(f_j(x_t) - \mu_j^+)$$

$$\mu_1^- = \frac{1}{|S^-|} \sum_{t \in S^-} f_1(x_t)$$

$$\mu_2^- = \frac{1}{|S^-|} \sum_{t \in S^-} f_2(x_t)$$

$$\Sigma_{i,j}^- = \frac{1}{|S^-|} \sum_{t \in S^-} (f_i(x_t) - \mu_i^-)(f_j(x_t) - \mu_j^-)$$

We let μ^+ be the vector $\langle \mu_1^+, \mu_2^+ \rangle$ and similarly for μ^- . Suppose that we model $P(y = 1)$ as $|S^+|/T$; model the conditional the conditional probability $P(\Phi(x)|y = 1)$ as a Gaussian with mean μ^+ and covariance Σ^+ ; and model $P(\Phi(x)|y = -1)$ similarly with mean μ^- and covariance Σ^- . Now suppose that we have a particular sample satisfying the following.

$$\begin{aligned}
|S^+| &= \frac{1}{2}T \\
\mu^+ &= \langle 0, 0 \rangle \\
\mu^- &= \langle 1, 0 \rangle \\
\langle z, w \rangle (\Sigma^+)^{-1} \langle z, w \rangle &= 2z^2 + w^2 \\
\langle z, w \rangle (\Sigma^-)^{-1} \langle z, w \rangle &= z^2 + 2w^2
\end{aligned}$$

Write the condition $P(y = 1|\Phi(x)) \geq P(y = -1|\Phi(x))$ as a quadratic condition on $f_1(x)$ and $f_2(x)$. Draw the decision boundry and label the region in which we have $P(y = 1|\Phi(x)) \geq P(y = -1|\Phi(x))$.

Problem 2. Again consider the sample from problem 1. Define S^+ and S^- as in problem 1.

a. Define a new feature space $\Psi(x) \in R^d$ (for some d) such that the quadratic decision boundry of problem 1 can be written as $\beta_D \cdot \Psi(x) \geq 0$. Give both the feature map Ψ and the vector β_D .

b. Define β^* as follows where Ψ is defined as in part a.

$$\beta^* = \operatorname{argmax}_{\beta} \sum_{t=1}^T \operatorname{sign}(y_t(\beta^T \Psi(x_t)))$$

Give an argument (physicists proof) that in the limit as $T \rightarrow \infty$ we should have that β^* is at least as good as β_D and usually better.

Problem 3. Here will show that the Gaussian kernel is indeed a kernel. Consider $x, y \in R^d$. The problem is to show that there exists a feature map Ψ , with $\Psi(x), \Psi(y) \in \ell_2$, such that $\exp(-\frac{1}{2}(x-y)^T \Sigma^{-1}(x-y)) = \Psi(x) \cdot \Psi(y)$.

a. Let $p(x)$ be an infinite series $a_0 + a_1x + a_2x^2 + a_3x^3, \dots$ with positive coefficients ($a_i \geq 0$) and that converges for all real values x . Show that, for any kernel K , we have that $p(K(x, y))$ is also a kernel. Hint: Let Φ be the

feature map for K and let $f_i(x)$ be the i th feature of x under the feature map Φ . Write $p(K(x, y))$ as an infinite sum of products of individual feature values. Then define a countable set of features (indexed by some countable set, but not necessarily the set of natural numbers) such that the infinite sum of products for $p(K(x, y))$ can be written as an inner product of feature vectors. Extra credit: explain where you need $a_i \geq 0$ and where you need that the series converges.

- b.** Show that $\exp(-\frac{1}{2}(x-y)^T \Sigma^{-1}(x-y))$ is equal to $\exp(-\frac{1}{2}x^T \Sigma^{-1}x) \exp(-\frac{1}{2}y^T \Sigma^{-1}y) \exp(x^T \Sigma^{-1}y)$.
- c.** Show that $\exp(-\frac{1}{2}x^T \Sigma^{-1}x) \exp(-\frac{1}{2}y^T \Sigma^{-1}y)$ is a kernel in x and y (Hint: you only need a single feature.)
- d.** Show that $x^T \Sigma^{-1}y$ is a kernel in x and y .
- c.** Use parts a. and d. to show that $\exp(x^T \Sigma^{-1}y)$ is a kernel in x and y .
- e.** Use the fact that a product of kernels is a kernel (proved in lecture) to show that the Gaussian kernel is a kernel.