SISE 9128: Introduction to Machine Learning

Alex Smola, RSISE ANU

Teaching Period
Problem Sheet — Week 2
October 8-19, 2001
The due date for these problems is Friday, October 19

Problem 10 (Convolutions and Random Variables)
Show that for two random variables ξ_1, ξ_2 with densities $p_1(\xi_1)$ and $p_2(\xi_2)$ the density of the random variable $\xi := \xi_1 + \xi_2$ is given by $p(\xi) = p_1 \circ p_2(\xi)$.

Problem 11 (Kernels, B_n-Splines, and Mercer's Condition)
In this problem we will introduce a new class of kernels. For this purpose denote by B_0 the indicator function on the interval $[-\frac{1}{2}, \frac{1}{2}]$, i.e.

$$B_0(x) = \begin{cases} 1 & \text{if } |x| \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}$$

Furthermore we introduce the splines B_n on $\mathbb{R}$ via $B_{n+1} := B_n \circ B_0$.

1. *Compute the splines B_1 and B_2 analytically.*
2. *Show that B_n is a spline of order n, i.e. it is piecewise polynomial up to order n.* **Hint:** *use induction, i.e. assume that it is true for B_n and show that it then holds for B_{n+1}.*
3. *Compute the Fourier transform of B_0. Why does it follow from this that $k(x, x') := B_0(x - x')$ is not a vaild kernel?*
4. *Show that the Fourier transform of B_n is given by $\tilde{B}_n = (2\pi)^{\frac{n}{2}} \left(\tilde{B}_0\right)^n$. Which $k(x, x') := B_n(x - x')$ is therefore a valid kernel?*
5. **Bonus question (difficult):** *Show that $p_n(x) := \frac{n+1}{12} B_n\left(\frac{n+1}{12} x\right)$ converges to a normal distribution with zero mean and unit variance.*

   Hint: *Use the result of Problem 10. Next show that B_n is the density corresponding to a sum of $n + 1$ random variables uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$. Finally, show that p_n has zero mean and unit variance and apply the central limit theorem (from second week) to prove the claim.*

Problem 12 (Radial Basis Function Kernels)
Denote by $k(\mathbf{x}, \mathbf{x}') := \kappa(\|\mathbf{x} - \mathbf{x}'\|)$ a radial basis function kernel.

1. *Show that for a strictly monotonically decreasing $\kappa : [0, \infty) \to \mathbb{R}$ the mapping into a feature space is neighbourhood preserving, i.e. that*

   $$d(\Phi(\mathbf{x}), \Phi(\mathbf{x}')) \leq d(\Phi(\mathbf{x}), \Phi(\mathbf{x}'')) \text{ is equivalent to } d(\mathbf{x}, \mathbf{x}') \leq d(\mathbf{x}, \mathbf{x}'')$$

2. *Show that for the kernels given below the feature map Φ maps all $\mathbf{x}$ onto the surface of a sphere, and more precisely, into an orthant of $90°$.*

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{x}'\|^2\right) \quad (1)$$

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{\sigma}\|\mathbf{x} - \mathbf{x}'\|\right) \quad (2)$$