Tim Paine
tkp2108

3 April 2014

Statistical Machine Learning

Problem 1

1. The one on the left encourages sparse estimates, as it looks for points in $\hat{\beta}$ that have either β_1 or β_2 but not both, whereas the one on the right does not encourage sparse elements (as clearly evidenced by the fact that x_4 and x_5 intersect with the cost function).
2. For the cost on the left, x_3 minimizes the cost, as it is the only point to intersect the constraint region. For cost on the right, two points x_3 and x_5 intersect the border of the constraint region, and one point x_4 lies within the constraint region. We are looking for a solution which minimizes the value on which we are constraining, so since x_4 lies within the region, its associated cost is less than the two points that lie on the border.

■

Problem 2

1. Show that for any positive $a \in \mathbb{R}$, $k(x, x') = ak_1(x, x')$ is a kernel.
We assume that $k_1(x, x') = \Phi(x)\Phi(x')$ for some Φ. Let $\hat{\Phi} = \sqrt{a}\Phi(x)$. Then clearly, $\Phi(\hat{x})\Phi(\hat{x'}) = ak_1(x, x')$, and therefore $ak_1(x, x')$ is also a kernel. The condition that a be positive comes from the fact that we need to be able to take the square root of a.

2. Show that $k(x, x') = k_1(x, x')k_2(x, x')$ is a kernel.
Again, for some Φ_1 and Φ_2, we have that $k_1(x, x') = \Phi_1(x)\Phi_1(x')$ and $k_2(x, x') = \Phi_2(x)\Phi_2(x')$. Let's look at $k_1 * k_2$. we have $k_1(x, x')k_2(x, x') = (\Phi_1(x)\Phi_1(x'))(\Phi_2(x)\Phi_2(x'))$. When we multiply this out, each set will form a sum, and we need to multiply the sums element-wise.
I.e., $(\Phi_1(x)\Phi_1(x'))(\Phi_2(x)\Phi_2(x')) = (\sum_i \phi_{1,i}(x)\phi_{1,i}(x'))(\sum_j \phi_{2,j}(x)\phi_{2,j}(x'))$. If we multiply these out element-wise, we can write this as $\sum_{i,j} \phi_{1,i}(x)\phi_{1,i}(x')\phi_{2,j}(x)\phi_{2,j}(x')$. If we define $\phi_{3,i,j}(x) = \phi_{1,i}(x)\phi_{2,j}(x)$, and $\Phi_3(x)\Phi_3(x') = \sum_{i,j} \phi_{3,i,j}(x)\phi_{3,i,j}(x')$ then it is becomes obvious that $k_1(x, x')k_2(x, x') = \Phi_3(x)\Phi_3(x')$, and therefore is also a kernel.

3. Show that for any positive $p \in \mathbb{Z}$, $k(x, x') = k_1(x, x')^p$ is a kernel.
Again, we assume that $k_1(x, x') = \Phi(x)\Phi(x')$ for some Φ. Let $\hat{\Phi} = \Phi^p(x)$. Then, because $p > 0, p \in \mathbb{Z}$, $\hat{\Phi}(x)\hat{\Phi}(x') = \Phi^p(x)\Phi^p(x') = k_1(x, x')^p$, so therefore $k_1(x, x')^p$ is also a kernel.

■