Chapter 9

stevenjin8

April 10, 2021

Comments and Proofs

4.4 Kernel PCA

It took me a while to understand this section. The idea is to leverage the Mercer property of kernels to map the data to a larger (potentially infinite) dimension feature space and to compute the principal components over said feature space. Given that, we first compute the Gram matrix:

$$\mathbf{K} = \mathbf{\Phi}\mathbf{\Phi}^T$$
$$k_{i,j} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$$

Using the eigenvalue/eigenvector trick presented earlier we find the formula for $\mathbf{V}_{kpca}$. Thus the kpca embedding of a data point $\mathbf{x}_*$ is $\phi(\mathbf{x}_*)\mathbf{\Phi}^T\mathbf{U}\mathbf{\Lambda}^{-\frac{1}{2}}$ (note that equation 14.40 is missing a transpose).

I still don't understand algorithm 14.2. Given some new data $\mathbf{X}_*$, the vectorized equation for $\tilde{\mathbf{K}}_*$ should be

$$\begin{aligned}\tilde{\mathbf{K}}_* &= (\mathbf{\Phi}_* - \frac{1}{N}\sum \phi_i)\mathbf{\Phi}^T\mathbf{U}_{:,1:z}\mathbf{\Lambda}_{:,1:z} \\ &= (\mathbf{K}_* - \mathbf{1}_{N_*}\overline{\mathbf{k}}^T - \overline{\mathbf{k}}_*\mathbf{1}_N^T + \overline{k}\mathbf{1}_{N_*}\mathbf{1}_N^T)\mathbf{U}_{:,1:z}\mathbf{\Lambda}_{:,1:z}\end{aligned}$$

where $\mathbf{K}_* = \mathbf{\Phi}_*\Phi^T$ contain the pairwise kernel between the new data and the training data; $\overline{\mathbf{k}}$ is the row-wise mean for $\mathbf{K}$; $\overline{\mathbf{k}}_*$ is the row-wise mean of $\mathbf{K}_*$; and $\overline{k}$ is the mean of all values in $\mathbf{K}$. There is an implementation of this in the notebooks folder.

Regardless, line 8 of the equation cannot be correct since both $\mathbf{O}_*$ and $\mathbf{K}_*$ are $N_* \times N$. Thus, three out of the four terms are not defined.

Something that I found really interesting is that we do not normalize the columns of $\mathbf{\Phi}$. It makes sense, however, the whole idea of KPCA is centered around the kernel function and dimensions in the feature space that have more extreme values are going to have a larger impact on the kernel.

Exercises

Exercise 1

a.