Kernel Methods

Homework 2

Luis Antonio Ortega Andrés
Antonio Coín Castro

March 3, 2021

1 TODO

Revisar cuaderno.

Aclarar que eliminamos las componentes no nulas.

Kernel PCA

Why do the projections onto the first two KPCA principal components look different for the Sklearn and our implementation? Is any of the two incorrect?

Both representations are correct and valid, what makes them different is the decisions made at normalizing the eigenvectors.

These vectors are returned with $\|\cdot\|_{\mathcal{L}_2} = 1$ in $\mathbb{R}^N$ by Numpy and Scipy via specialized functions such as `eig` and `eigh` (version for symmetric cases). However, the kernel method needs these vectors to be unitary in the subjacent Hilbert space, not $\mathbb{R}^N$. This can be achieved by ensuring that their $\mathcal{L}_2$-norm in $\mathbb{R}^N$ equals the inverse of the square root of their corresponding eigenvalues $1/\sqrt{\lambda}$. The procedure is then to divide each eigenvector by this quantity, and it is in this step that generates discrepancy, as one may divide by $-1/\sqrt{\lambda}$ getting the same $\mathcal{L}_2$-norm.

Sklearn's implementation uses an auxiliary function `svd_flip` that deterministically assigns a sign for each normalized eigenvector (reference).

We have implemented this option as a parameter for our function in order to reproduce Sklearn's output perfectly. In our case, not using this function leads to a reflection of the second principal component in the Kernel PCA function. Linear PCA does not show this behavior by pure coincidence.

Vary the parameters of the kernel and comment on the behavior of the projections onto the first two KPCA components for the different values considered (e.g. $\gamma \in \{0.02, 0.2, 2.0, 20.0, 200.0, 2000.0\}$*. In particular,*

1. *What is the behaviour in the limit in which the width of the kernel approaches* ∞*. Explain why one should expect such behavior.*
2. *What is the behaviour in the limit in which the width of the kernel approaches* 0*. Explain why one should expect such behavior.*

First of all, to study the limit behaviour of the projections we are using the explicit kernel formula:

$$\mathcal{K}(x, y) = Ae^{-\frac{\|x-y\|^2}{2\omega^2}},$$

where A is the output variance and ω is the kernel's width and γ is inversely proportional to ω . Using this formula, and given two fixed input values x and y,

$$\lim_{\gamma \to 0^+} = \lim_{\omega \to \infty} \mathcal{K}(x, y) = A.$$

$$\lim_{\gamma \to \infty} = \lim_{\omega \to 0^+} \mathcal{K}(x, y) = 0.$$

Using this theoretical results, increasing the value of gamma result in projecting all the datapoints into 0, and decreasing it

Nyström approximation

Comment on the values of the error for the different approximations, and their dependence with the number of sampled features.

(Extra point) Determine de dependence of the mean error with the number of features for the different random feature models. Provide an explanation of this behavior.

2 CV

DUDA:¿adaptamos nuestro código para poder usarlo con sklearn? (heredar de BaseEstimator, pasar los parámetros en el constructor, que el fit sea fit de verdad, etc...)

References

Formato Chicago de todos los trabajos consultados + sklearn.

Appendix: code developed

DUDA: ¿Qué código hay que poner aquí?