

Kernel Methods

Homework 2

Luis Antonio Ortega Andrés
Antonio Coín Castro

March 2, 2021

1 Kernel PCA

2. 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?

Los vectores α tienen norma (l2) igual a 1 en \mathbb{R}^N , y se demuestra que si conseguimos que tengan norma (l2) $1/\sqrt{\lambda_i}$ entonces los autovectores V_i definidos por α_i en el espacio H tendrán norma-H igual a 1. Esto se reduce a dividir $\alpha_i/\sqrt{\lambda_i}$.

Las dos representaciones mostradas son correctas, solo que una está usando los vectores α cambiados de signo (siempre hay dos posibilidades). Sklearn usa la función `svd_flip` para unificar la elección de signo (<https://github.com/scikit-learn/scikit-learn/blob/15c2c72e27c6ea18566f4e786506c7a3aef8a5de/sklearn/utils/extmath.py#L504>). El efecto visual es una simetría, ya que en este caso cambia solo el autovector correspondiente a la segunda componente principal.

En PCA lineal no cambia por casualidad.

DUDA: ¿qué hacemos con las componentes/autovalores nulos? ¿Devolvemos columnas con todo 0s? ¿O las eliminamos antes? ¿O no hacemos nada?

3. 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,

a) What is the behaviour in the limit in which the width of the kernel approaches ∞ . Explain why one should expect such behavior.

b) What is the behaviour in the limit in which the width of the kernel approaches 0. Explain why one should expect such behavior.

2 Nyström approximation

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

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

DUDA: Qué hay que hacer exactamente con `Xprime` en Nystrom. Para qué sirve la función `approximate_kernel_matrix`? ¿No sería mejor que aceptara `X` y `X_prime`?

DUDA: ¿Hay que entregar lo de mnist?

3 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?