— *Solution notes* —

Fifth practice exercises in Machine learning 1 – 2023 – Paper 1

1 Principal component analysis (September)

Suppose we have a data set $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ of D-dimensional vectors, which have a zero mean for each dimension. Assume we perform a complete eigenvalue decomposition of the empirical covariance matrix $\mathbf{S} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$. You are interested in only a single projection of your data such that the variance of this projection is maximized. Let $\mathbf{u}_i$ be the direction vector of a particular projection. Assume that $\mathbf{u}_i^T\mathbf{u}_i = 1$.

(*a*) What is the projection z_{ni} of a given point $\mathbf{x}_n$ under the particular vector $\mathbf{u}_i$?

Answer:

The projection of the vector $\mathbf{x}_n$ over the vector $\mathbf{u_i}$ is given by:

$$z_{ni} = \mathbf{u_i}^T\mathbf{x}_n$$

(*b*) What is the empirical mean of the projection z_i across all points $\mathbf{x}_n$?

Answer: The empirical mean:

$$E[z_i] = \frac{1}{N}\sum_{n=1}^{N}\mathbf{u}_i^T\mathbf{x}_n = \mathbf{u}_i^T\left(\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n\right) = 0$$

(*c*) What is the empirical variance of the projection z_i? Provide your answer in terms of the empirical covariance matrix $\mathbf{S}$

Answer:

The variance of a variable $\mathbf{z}$ is defined as:

$$V[z_i] = E[(z_i - \bar{z}_i)^2]$$

Hence, in our case:

$$V[z_i] = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{u}_i^T\mathbf{x}_n)(\mathbf{u}_i^T\mathbf{x}_n)^T$$

$$= \frac{1}{N}\sum_{n=1}^{N}\mathbf{u}_i^T\mathbf{x}_n\mathbf{x}_n^T\mathbf{u}_i$$

$$= \mathbf{u}_i^T\mathbf{S}\mathbf{u}_i$$

(*d*) Replace $\mathbf{S}$ with its eigenvalue decomposition and simplify the aforementioned expression. What is the variance now?

Answer:

$$V[z_i] = \mathbf{u}_i^T \mathbf{S} \mathbf{u}_i = \mathbf{u}_i^T \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T \mathbf{u}_i$$

Since $\mathbf{u}_i^T \mathbf{u}_i = 1$ and $\mathbf{u}_i^T \mathbf{u}_j = 0$ then we can re-write:

$$= \mathbf{e}_i^T \Lambda \mathbf{e}_i = \lambda_i$$

---

with $\mathbf{e}_i$ to be a vector with zeros except the position with index i.

(e) Suppose that you are interested in reducing the dimensionality from D to K, such that 99% of the variance is maintained. How can you select an appropriate K?

Answer: We need to sort eigenvalues in descending order. Then by picking K largest eigenvalues using the following formula:

$$\frac{\sum_{i=1}^{K-1} \lambda_i}{\sum_{i=1}^{D} \lambda_i} < 0.99 \leq \frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{D} \lambda_i}$$
