61. Kernel PCA. Spring 2014

Note Title 5/11/2014

Now we show that the kernel trick can be applied to PCA.

In this lecture we set the means to zero (e.g. $\sum_{i=1}^N x_i = 0$); this is easy to arrange (by subtraction). $\sum_{i=1}^N \varphi(x_i) = 0$

The output of PCA is the projection of the data $\{x_i : i = 1 \dots N\}$ onto the subspace defined by $\{e_1, \dots, e_M\}$, i.e. the coefficients $(x_i \cdot e_1, x_i \cdot e_2, \dots, x_i \cdot e_M)$. $i = 1 \dots M$

This depends on dot products, which suggests we can use the kernel trick if we replace x by $\varphi(x)$.

But what about the e's? They are eigenvectors of the correlation function. How do they change if we replace x by $\varphi(x)$?

To understand this, we need another way to compute the eigenvectors of $C = \frac{1}{N} \sum_{i=1}^N x_i x_i^T$

Claim. The eigenvectors with non-zero eigenvalues can be expressed in form $e = \sum_{i=1}^N \alpha_i x_i$. The eigenvectors with zero eigenvalues are of form e, s.t. $x_i \cdot e = 0$, $i = 1 \dots N$.

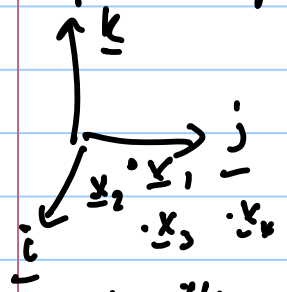
Proof. Suppose $x_i \cdot e = 0$, $i = 1 \dots N$

then $Ce = \frac{1}{N} \sum_{i=1}^N x_i (x_i \cdot e) = 0$, hence e is an eigenvector with zero eigenvalue. This proves the last sentence.

The remaining eigenvectors must be orthogonal to the zero eigenvectors. Hence they are of form $\sum_{i=1}^N \alpha_i e_i$.

(2) Intuition: Suppose we have data in 3-dim space spanned by axes $\underline{i}, \underline{j}, \underline{k}$. Suppose all the data lies in the plane spanned by $\underline{i}, \underline{j}$

Note Title 11/19/2006



i.e. $\underline{x}_1 = x_1 \underline{i} + y_1 \underline{j}$

$\underline{x}_2 = x_2 \underline{i} + y_2 \underline{j}$

$\underline{x}_N = x_N \underline{i} + y_N \underline{j}$

Then all vectors in the $\underline{i}, \underline{j}$ plane can be written as $\underline{y} = \sum_{i=1}^{N} \alpha_i \underline{x}_i$.

The eigenvectors of the correlation function lie in the $\underline{i}, \underline{j}$ plane, except for the zero eigenvalue with eigenvector $\underline{e} = \underline{k}$

Now replace $\underline{x}$ by $\underline{\varphi}(\underline{x})$

$$\rightarrow \underline{\underline{C}} = \frac{1}{N} \sum_{n=1}^{N} \underline{\varphi}(\underline{x}_n)\, \underline{\varphi}^T(\underline{x}_n)$$

All non-zero eigenvectors $\underline{e}$ of $\underline{\underline{C}}$ are of form

$$\underline{e} = \sum_{j=1}^{N} \alpha_j \underline{\varphi}(\underline{x}_j), \quad \text{for some } \{\alpha_j\},$$

Substituting: $\underline{\underline{C}}\, \underline{e} = \lambda \underline{e}$

$$\rightarrow \frac{1}{N} \sum_{n=1}^{N} \underline{\varphi}(\underline{x}_n) \{\underline{\varphi}(\underline{x}_n) \cdot \underline{e}\} = \lambda \underline{e}$$

$$\rightarrow \frac{1}{N} \sum_{n=1}^{N} \underline{\varphi}(\underline{x}_n) \sum_{j=1}^{N} \alpha_j \{\underline{\varphi}(\underline{x}_n) \cdot \underline{\varphi}(\underline{x}_j)\} = \lambda \sum_{j=1}^{N} \alpha_j \underline{\varphi}(\underline{x}_j)$$

Equating coeffs of $\underline{\varphi}(\underline{x}_j)$ gives new eigenvalue equations

$$\frac{1}{m} \sum_j K(\underline{x}_k, \underline{x}_j) \alpha_j = \lambda \alpha_k \quad \| \text{Index } \lambda^\mu \; \alpha_k^\mu$$

(3)

$$\frac{1}{N}\sum_j K(\underline{x}_k, \underline{x}_j)\alpha_j^\mu = \lambda^\mu \alpha_k^\mu \qquad \mu = 1, \ldots, p.$$

Solving this, gives us the eigenvectors

$$\underline{e}^\mu = \sum_{j=1}^{N} \alpha_j^\mu \, \underline{\phi}(\underline{x}_j), \quad \text{eigenvalue } \lambda^\mu.$$

(depends on ϕ)

But the projections $\underline{e}^\mu \cdot \underline{\phi}(\underline{x})$ of the data are

$$\underline{e}^\mu \cdot \underline{\phi}(\underline{x}) = \sum_{j=1}^{N} \alpha_j^\mu K(\underline{x}_j, \underline{x})$$

which is independent of ϕ.
(depends only on K).

Hence:

the projection of the data onto the eigenvectors requires only knowing the kernel $K(\underline{x}_i, \underline{x}_j)$ (i.e. not knowing ϕ)

Knowledge of the kernel is used twice:
(1) to compute the $\{\alpha_j^\mu\}$,
(2) to compute the projections $\underline{e}^\mu \cdot \underline{\phi}(\underline{x})$. //