Self-adjoint Kernels

Recap

Every H-S kernel has a singular value expansion

$$K(x,y) = \sum_{j=1}^{\infty} \sigma_j u_j(x) \overline{v_j(y)}$$

which converges in $L^2(\mathbb{R}^d \times \mathbb{R}^d)$.

$\Rightarrow$ Best rank-n approx in $\|\cdot\|$ and $\|\cdot\|_{HS}$ is

$$K_n(x,y) = \sum_{j=1}^{n} \sigma_j u_j(x) v_j(y)$$

$\Rightarrow$ $T = \lambda I + K$ satisfies Fredholm alternative:

a) T^{-1} exists and $\|T^{-1}\| < \infty$, or

b) $N(T)$ is nontrivial and $Tu = f$ has solutions IFF $\langle f, v \rangle = 0$ for all $v \in N(T^*)$.

$\Rightarrow$ K is a compact operator, meaning it can be approximated in norm (uniformly on all of $L^2(\mathbb{R}^d)$) by finite-rank operators.

Self-adjoint Kernels

If $K(x,y) = \overline{K(y,x)}$, then $K = K^*$ is self-adjoint.

The eigenvectors/values of K are solutions of

$$K u = \overset{\text{eigenvalue}}{\lambda} \underset{\text{eigenvector}}{u}$$

$\Rightarrow$ Every self-adjoint compact operator $K : H \to H$ has an ONB of eigenvectors

$\Rightarrow$ Eigenvalues are real and $\lambda_k \to 0$ as $k \to \infty$.

$\Rightarrow$ Eigenvector expansion of kernel is

$$K(x,y) = \sum_{j=1}^{\infty} \lambda_j u_j(x) \overline{u_j(y)}$$

which coincides with SVE when $\lambda_j \geq 0$, for $j = 1, 2, 3, \ldots$, i.e., K is semidefinite.

Example: Principle Component Analysis (PCA)

Suppose that $x_1, x_2, x_3, \dots, x_m \in \mathbb{R}^n$ are a sequence of identically and independently distributed random vectors with mean

$$\mu = \mathbb{E}[X] = \big(\underbrace{\mathbb{E}[x^{(1)}], \dots, \mathbb{E}[x^{(n)}]}_{\text{expected value of each component}}\big)^T$$

and covariance matrix

$$C_{ij} = \mathbb{E}\left[(x^{(i)}-\mu^{(i)})(x^{(j)}-\mu^{(j)})\right] \Rightarrow C = \mathbb{E}\big[\underbrace{(x-\mu)(x-\mu)^T}_{\text{expected value of rank-1 matrix}}\big]$$

Intuition: think of vectors $x_1, \dots, x_m$ as data collected from m identical, independent experiments. Each entry of x_j represents a particular "feature" measured during the j^{th} experiment and the same "features" are measured across all m experiments.

In practice, one can use empirical estimates of the mean & covariance:

$$\tilde{\mu} = \frac{1}{m} \sum_{j=1}^{m} u_j \qquad \left(\text{average features across all experiments}\right)$$

$$\tilde{C} = \frac{1}{m-1} \sum_{j=1}^{m} (x_j - \mu)(x_j - \mu)^T$$

Note that in general, the cov. matrix is <u>not</u> rank 1.

How should we analyze the data? Principle Component Analysis follows 2 guiding principles.

(i) ⇒ Find a new coordinate system such that the leading coordinate directions captures as much variance in the data as possible.

(ii) ⇒ Choose the coordinate system so that the features "decouple," i.e., the features "look like" independent variables (more precisely, diff. feat. have zero covariance).

The constraint is that the new coordinate system should be a rotation/reflection of the old coordinate system (features).

$$\text{new features} \rightarrow \quad y_j = \underset{\substack{\uparrow \\ \text{orthogonal} \\ \text{transform}}}{U^T}(x_j - \mu) \quad \leftarrow \text{old features}$$

Guided by principles (i)-(ii), how should we choose the orthogonal transformation U?

(ii) Can we choose U so that the random variable $y = U^T(x-\mu)$ has diagonal covariance?

$$\begin{aligned} \underbrace{\mathbb{E}[yy^T]}_{\substack{\text{cov. of} \\ \text{r.v. } y}} &= \mathbb{E}[U^T(x-\mu)(x-\mu)^T U] \\ &= U^T \mathbb{E}[(x-\mu)(x-\mu)^T] U \\ &= U^T C U \end{aligned}$$

To diagonalize $\text{cov}(y) = \mathbb{E}[yy^T]$, we choose U to be the eigenvector matrix of $C \Rightarrow C = U \Lambda U^T$

This is possible b/c covariance matrices are symmetric positive semi-definite matrices!

$$C_{ij} = \mathbb{E}[(x-\mu)(x-\mu)^T] = C_{ji} \quad \text{(symmetric)}$$

$$\begin{aligned} v^T C v &= \mathbb{E}[v^T(x-\mu)(x-\mu)^T v] \\ &= \mathbb{E}[(v^T(x-\mu))^2] \geq 0 \quad \text{(positive semi-def.)} \end{aligned}$$

$\Rightarrow U$ is an orthogonal matrix

$\Rightarrow \Lambda$ is diagonal w/ real non-negative entries

Therefore, $y_1, \ldots, y_m$ are distributed w/ mean zero and diagonal covariance matrix:

$$\mathbb{E}[yy^T] = U^T C U = \Lambda \quad \checkmark$$

$$\begin{aligned} \mathbb{E}[y] &= \mathbb{E}[U^T(x-\mu)] \\ &= U^T \mathbb{E}[x-\mu] \\ &= 0 \quad \checkmark \end{aligned}$$