Recitation 3

Kernels, Least Squares, and Complex Numbers

Brett Bernstein

CDS at NYU

February 11, 2019

Questions

1. Let $X \in \mathbb{R}^{p\times n}$ denote a matrix whose **rows** are datapoints $\vec{x}_1^T, \ldots, \vec{x}_n^T \in \mathbb{R}^p$ with $p \leq n$. Suppose you are only given access to $G = XX^T$. How would you compute the first $k < p$ principal components of $\vec{x}_i$, for $i = 1, \ldots, n$?
2. Generalizing the previous example, suppose there is a (possibly unknown) mapping $\Phi : \mathbb{R}^p \to \mathcal{H}$ where $\mathcal{H}$ is some Hilbert space. Suppose we are given a known, relatively easy to compute function $K : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ (called a *kernel*) such that $K(\vec{x}, \vec{y}) = \langle \Phi(\vec{x}), \Phi(\vec{y})\rangle_{\mathcal{H}}$.
   1. Given a dataset $\vec{x}_1, \ldots, \vec{x}_n \in \mathbb{R}^p$, if we compute the matrix $G \in \mathbb{R}^{n\times n}$ where $G_{ij} = K(\vec{x}_i, \vec{x}_j)$, what property of matrices must G have?
   2. Suggest a method for using K to perform a modified version of PCA (called Kernel PCA).
   3. One such kernel (the RBF or Gaussian kernel) is given by $K(\vec{x}, \vec{y}) = \exp(-\|\vec{x} - \vec{y}\|^2/\sigma^2)$. What does the fact that K is always positive-valued say about $\mathcal{H}$?

Questions

1. Let $X \in \mathbb{R}^{n\times p}$, $\beta \in \mathbb{R}^p$, $y \in \mathbb{R}^n$, $\lambda > 0$, and $M \in \mathbb{R}^{m\times p}$ has full column rank. What is the solution to $\arg\min_\beta \|X\beta - y\|_2^2 + \lambda\|M\beta\|_2^2$?
2. Suppose you are given data $\vec{y} = X\vec{\beta} + \vec{z}$ (all variables deterministic; $\vec{\beta}, \vec{z}$ unknown) and compute the least squares estimator $\hat{\vec{\beta}}$ for $\vec{\beta}$. Assuming $\|\vec{z}\|_2 = \eta$ is fixed, and X has full column rank, what direction for $\vec{z}$ produces the largest error $\|\hat{\vec{\beta}} - \vec{\beta}\|_2$, and how much is that error?
3. Let $\vec{\beta}_{\text{ridge}}$ denote the ridge regression estimator which minimizes $\vec{\beta}_{\text{ridge}} := \arg\min_{\vec{\beta}} \|X\vec{\beta} - \vec{y}\|_2^2 + \lambda\|\vec{\beta}\|_2^2$. Show how to compute $\vec{\beta}_{\text{ridge}}^T x$ on a new test point x using only $\vec{y}$, λ, and XX^T. What does this enable?

Questions

1. Let $(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R} \times \mathbb{R}$ be a given dataset. And consider the objective $L(h) = \frac{1}{n}\sum_{i=1}^{n}(h(x_i) - y_i)^2$, where $h : \mathbb{R} \to \mathbb{R}$. In each of the following, show how to solve the problem using linear least squares.
    1. Minimize $L(h)$ for $h \in \mathcal{F}$ where $\mathcal{F} = \{h(x) = a + bx + cx^2 + dx^3 : a, b, c, d \in \mathbb{R}\}$.
    2. Minimize $L(h)$ for $h \in \mathcal{F}$ where $\mathcal{F}$ is the set of piecewise cubic polynomials with knot locations $\tau_1 < \cdots < \tau_k$. That is, there are a_j, b_j, c_j, d_j for $i = 1, \ldots, k+1$ such that $h(t) = a_j + b_j t + c_j t^2 + d_j t^3$ for $\tau_{i-1} \leq h(t) < \tau_j$ where it is assumed that all of the data points lie in the interval (τ_0, τ_{k+1}).
    3. Repeat the previous part in each of the following cases:
        1. $h \in \mathcal{F} \cap \mathcal{C}^2$, i.e., h is a twice continuously differentiable cubic spline.
        2. $h \in \mathcal{F} \cap \mathcal{C}^2$, and the function is affine for $t < \tau_1$ and $t > \tau_k$ (called a natural cubic spline).

Questions

1. Compute $1 + e^{2\pi i/n} + e^{2\pi i/n} + \cdots + e^{2\pi(n-1)i/n}$ where $n > 1$.
2. Prove that Cauchy-Schwarz holds in a complex inner product space:

$$|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\|\|\vec{y}\|.$$

3. Show that $|r_1 e^{it} - r_2 e^{is}| \geq |r_1 - r_2|$ for all $r_1, r_2 > 0$ and $t, s \in \mathbb{R}$.