Themen des statistischen maschinellen Lernens
Wintersemester 2016/17
Humboldt-Universität zu Berlin

Dr. Martin Wahl

Sheet 2

1. A linear map $P : \mathbb{R}^p \to \mathbb{R}^p$ is called orthogonal projection if (i) P is idempotent, i.e. $P^2 = P$ and (ii) P is self-adjoint, i.e. $\langle Pu, v\rangle = \langle u, Pv\rangle$ for all $u, v \in \mathbb{R}^p$.

   (a) Let P be an orthogonal projection and $\operatorname{Im} P = \{Pu : u \in \mathbb{R}^p\}$ be the image of P. Show that $\langle u - Pu, v\rangle = 0$ for all $u \in \mathbb{R}^p$ and $v \in \operatorname{Im} P$. Deduce that $\|u - Pu\|^2 \leq \|u - v\|^2$ for all $u \in \mathbb{R}^p$ and $v \in \operatorname{Im} P$.

   (b) Given a d-dimensional subspace V of $\mathbb{R}^p$, there is a unique orthogonal projection P with $\operatorname{Im} P = V$. If $v_1, \ldots, v_d$ is an orthonormal basis of V, it can be written (in matrix form) as $P = \sum_{j\leq d} v_j v_j^T = (v_1 \cdots v_d)(v_1 \cdots v_d)^T$.

2. (Best affine approximation). Let $X_1, \ldots, X_n \in \mathbb{R}^p$. Consider the following minimisation problem
$$\min_{\mu, z_i, V} \sum_{i=1}^n \|X_i - \mu - V z_i\|^2, \tag{0.1}$$
where the minimum is taken over all $\mu \in \mathbb{R}^p$, $z_1, \ldots, z_n \in \mathbb{R}^d$ and $V \in \mathbb{R}^{p\times d}$ whose columns form an orthonormal system in $\mathbb{R}^p$. Let
$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \quad\text{and}\quad \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T.$$
By the spectral theorem, we have $\hat{\Sigma} = \sum_{j=1}^p \hat{\lambda}_j \hat{u}_j \hat{u}_j^T$ with $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$ and $\hat{u}_1, \ldots, \hat{u}_p$ orthonormal basis of $\mathbb{R}^p$. Show that the minimum in (0.1) is attained for $\mu = \bar{X}$, $V = (\hat{u}_1, \ldots, \hat{u}_d)$ and $z_i = V^T(X_i - \bar{X})$.

3. Let $X_1, \ldots, X_n \in \mathbb{R}^p$ and $\hat{\Sigma} = (1/n)\sum_{i=1}^n X_i X_i^T$. By the spectral theorem, we have $\hat{\Sigma} = \sum_{j=1}^p \hat{\lambda}_j \hat{u}_j \hat{u}_j^T$ with $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$ and $\hat{u}_1, \ldots, \hat{u}_p$ orthonormal basis of $\mathbb{R}^p$. Show that for $j = 1, \ldots, p$,
$$\hat{u}_j \in \underset{\substack{\|v\|=1 \\ v\perp \hat{u}_1,\ldots,\hat{u}_{j-1}}}{\operatorname{argmax}} \langle \hat{\Sigma} v, v\rangle = \underset{\substack{\|v\|=1 \\ v\perp \hat{u}_1,\ldots,\hat{u}_{j-1}}}{\operatorname{argmax}} \frac{1}{n}\sum_{i=1}^n \langle X_i, v\rangle^2.$$

4. (Reproducing kernel Hilbert space). A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a positive definite kernel if it is symmetric ($k(x, y) = k(y, x)$ for all $x, y \in \mathcal{X}$) and if for any $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \mathcal{X}$ and $a_1, \ldots, a_n \in \mathbb{R}$ we have
$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j k(x_i, x_j) \geq 0.$$

The goal of this exercise is to show that there is a Hilbert space $\mathcal{F}$ of functions on $\mathcal{X}$ such that (i) $k(x,\cdot) \in \mathcal{F}$ for all $x \in \mathcal{X}$ and (ii) $f(x) = \langle f, k(x,\cdot)\rangle_{\mathcal{F}}$ for all $x \in \mathcal{X}$ and $f \in \mathcal{F}$.

(a) Let

$$\mathcal{F}_0 = \left\{ f : \mathcal{X} \to \mathbb{R} : f = \sum_{i=1}^{n} a_i k(x_i, \cdot), n \in \mathbb{N}, x_i \in \mathcal{X}, a_i \in \mathbb{R}, i = 1, \dots, n \right\}.$$

For $f = \sum_{i=1}^{n} a_i k(x_i, \cdot)$ and $g = \sum_{j=1}^{m} b_j k(y_j, \cdot)$, define

$$\langle f, g\rangle_{\mathcal{F}_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j k(x_i, y_j).$$

Show that $(\mathcal{F}_0, \langle\cdot,\cdot\rangle_{\mathcal{F}_0})$ is an inner product space and that $f(x) = \langle f, k(x,\cdot)\rangle_{\mathcal{F}_0}$ for all $x \in \mathcal{X}$ and $f \in \mathcal{F}_0$.

(b) Show that that for any Cauchy sequence (f_n) in $\mathcal{F}_0$, the sequence $(f_n(x))$ converges in $\mathbb{R}$ for all $x \in \mathcal{X}$. (*Hint*: Use that $|f(x)| \le \sqrt{k(x,x)}\|f\|_{\mathcal{F}_0}$.)

(c) Show that for any Cauchy sequence (f_n) in $\mathcal{F}_0$ which converges pointwise to 0, we have $\|f_n\|_{\mathcal{F}_0} \to 0$.

(d) (Completion). Let $\mathcal{F}$ be the class of functions $f : \mathcal{X} \to \mathbb{R}$ which are pointwise limits of Cauchy sequences (f_n) in $\mathcal{F}_0$. For $f, g \in \mathcal{F}$, define

$$\langle f, g\rangle_{\mathcal{F}} = \lim_{n\to\infty} \langle f_n, g_n\rangle_{\mathcal{F}_0},$$

where (f_n) and (g_n) are Cauchy sequences converging pointwise to f and g, respectively. Prove that $\langle\cdot,\cdot\rangle_{\mathcal{F}}$ is a (well-defined) inner product on $\mathcal{F}$. Show that $(\mathcal{F}, \langle\cdot,\cdot\rangle_{\mathcal{F}})$ is a Hilbert space which contains $\mathcal{F}_0$ as a dense subspace. Deduce that $\mathcal{F}$ is a Hilbert space which satisfies the above properties (i) and (ii).