DS-GA 1003: Homework 4
Kernels, Duals, and Trees

Due on Tuesday, March 22, 2016

Professor David Ronsenberg

See complete code at: *git@github.com:cryanzpj/1003.git*

Yuhao Zhao
Yz3085

2 Positive Semidefinite Matrices

– 2.1.

Let $A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, A^T A = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and A is not symmetric.

– 2.2.

Since M is psd, we assume M is real and symmetric, thus by Spectral Theorem, we have

$$M = Q\Sigma Q^T$$

where Q is an orthogonal matrix ($Q^T = Q^{-1}$), Σ is diagonal.

$$\Sigma = Q^{-1}M(Q^T)^{-1} = Q^T M Q = \begin{pmatrix} q_1^T M q_1 & q_1^T M q_2 & \cdots & q_1^T M q_n \\ q_2^T M q_1 & q_2^T M q_2 & \cdots & q_2^T M q_n \\ \vdots & \vdots & \cdots & \vdots \\ q_d^T M q_1 & q_d^T M q_2 & \cdots & q_d^T M q_n \end{pmatrix}.$$

Since M is psd, $q_i^T M q_i \geq 0$, the diagonals of Σ are eigenvalues of M and are non-negative.

– 2.3.

i). If we have M $= BB^T$ for some B, for $\forall v \in \Re^n$

$$v^T M v = v^T B B^T v = (B^T v)^T (B^T v) = ||B^T v|| \geq 0$$

Therefore, M is psd
ii). If we know M is psd, by spectral theorem

$$M = Q\Sigma Q^T = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^T = Q\Sigma^{\frac{1}{2}}(Q\Sigma^{\frac{1}{2}})^T = BB^T$$

where $B = Q\Sigma^{\frac{1}{2}}$
$\Sigma^{\frac{1}{2}}$ is a diagonal matrix whose diagonal equals the sqrt root of $diag(\Sigma)$

This proves a symmetric matrix M can be expressed as $M = BB^T$ iff M is psd

3 Positive Definite Matrices

– 3.1.

M is pd, by Spectral Theorem,

$$M = Q\Sigma Q^T$$

$$\Sigma = Q^{-1}M(Q^T)^{-1} = Q^T M Q = \begin{pmatrix} q_1^T M q_1 & q_1^T M q_2 & \cdots & q_1^T M q_n \\ q_2^T M q_1 & q_2^T M q_2 & \cdots & q_2^T M q_n \\ \vdots & \vdots & \cdots & \vdots \\ q_d^T M q_1 & q_d^T M q_2 & \cdots & q_d^T M q_n \end{pmatrix}.$$

Since M is pd, $q_i^T M q_i > 0$, the diagonals of Σ are eigenvalues of M and are positive .

– **3.2.**

since M is positive definite, $M = Q\Sigma Q^T$

$$Q\Sigma Q^T M = Q\Sigma^{-1}Q^T Q\Sigma Q^T$$

Q is an orthogonal matrix, $Q^T Q = I$

$$RHS = Q\Sigma^{-1}\Sigma Q^T = QQ^T = I$$

Therefore, $Q\Sigma Q^T$ is the inverse of M

– **3.3.**

M is psd and symmetric, for $\forall v \in \Re^n, v \neq \vec{0},$ and $\lambda > 0$

$$v^T(M + \lambda I)v = v^T M v + \lambda v^T v > 0$$

since $v^T M v \geq 0, \lambda v^T v > 0$. Therefore, $v^T(M + \lambda I)v$ is positive definite.
To show, $M + \lambda I$ is symmetric, we know that $\forall i \neq j, (M + \lambda I)_{i,j} = M_{i,j} = M_{j,i} = (M + \lambda I)_{j,i}$. Thus $M + \lambda I$ is also symmetric.

let $v_1, ..., v_n, \lambda_1, ..., \lambda_n$ be the n eigenvalues and eigenvectors of M

$$(M + \lambda I)v_i = Mv_i + \lambda v_i = \lambda_i v_i + \lambda v_i = (\lambda_i + \lambda)v_i$$

Therefore, v_i is also a eigenvector of $M + \lambda I$ with corresponding eigenvalue equals to $(\lambda_i + \lambda)$.
$M + \lambda I = Q\Sigma Q^T, Q = \{v_1, ..., v_n\}, \Sigma_{i,i} = \lambda_i + \lambda$, Then we have

$$(M + \lambda I)^{-1} = (Q^T)^{-1}\Sigma^{-1}Q^{-1} = Q\Sigma^{-1}Q^T = \sum_{i=1}^{n} \frac{1}{\lambda_i + \lambda} v_i v_i^T$$

– **3.4.**

M is symmetric psd and N is symmetric pd, $\forall v \in \Re^n, v \neq \vec{0}$

$$v^T(M + N)v = v^T M v + v^T N v$$

we know $v^T M v \geq 0, v^T N v > 0$

$$v^T(M + N)v > 0$$

This shows $M + N$ is positive definite.
To show $M + N$ is symmetric, $\forall i \neq j, (M + N)_{i,j} = M_{i,j} + N_{i,j} = M_{j,i} + N_{j,i} = (M + N)_{j,i}$, Thus $M + N$ is also symmetric. From 3.2 we know that positive definite matrix has inverse. Therefore, M+N is invertible.

4 Kernel Matrices

$$K = XX^T = \begin{pmatrix} x_1^T x_1 & \cdots & x_1^T x_m \\ \vdots & \vdots & \vdots \\ x_m^T x_1 & \cdots & x_m^T x_m \end{pmatrix}$$

$d(x_i, x_j) = ||x_i - x_j|| = \sqrt{(x_i - x_j) \cdot (x_i - x_j)} = \sqrt{x_i \cdot x_i + x_j \cdot x_j - 2x_i \cdot x_j} = \sqrt{K_{i,i} + K_{j,j} - 2K_{i,j}}$
Therefore , knowing K is equivalent to knowing the set of pairwise distance of vectors in S.

5 Kernel Ridge Regression

– 5.1.

Since

$$J(w) = ||Xw - y|| + \lambda ||w^2|| \tag{1}$$

$$\frac{\partial J}{\partial w} = 2X^T(Xw - y) + 2\lambda wI = 0 \tag{2}$$

we have

$$X^T Xw - X^T y + \lambda wI = (X^T X + \lambda I)w - X^T y = 0 \tag{3}$$

$$w^* = (X^T X + \lambda I)^{-1} X^T y \tag{4}$$

XX^T is positive semidefinite and $\lambda > 0$, by 3.3, $XX^T + \lambda I$ is positive definite, thus invertible.

– 5.2.

Since $X^T Xw + \lambda Iw = X^T y, w = \frac{1}{\lambda}(X^T y - X^T Xw) = X^T \frac{1}{\lambda}(y - Xw)$
Thus $w = X^T \alpha$, where $\alpha = \frac{1}{\lambda}(y - Xw)$

– 5.3.

Since $w = X^T \alpha = \sum_1^n \alpha_i x_i$, w is a linear combination of data vectors

– 5.4.

since $w = X^T \alpha$ and $X^T Xw + \lambda Iw = X^T y$

$$X^T XX^T \alpha + \lambda IX^T \alpha = X^T y \tag{5}$$

$$X^T(XX^T + \lambda I)\alpha = X^T y \tag{6}$$

Therefore $\alpha = (XX^T + \lambda I)^{-1} y$

– 5.5.

Since $w = X^T \alpha = X^T (XX^T + \lambda I)^{-1} y, XX^T = K$

$$Xw = XX^T(XX^T + \lambda I)^{-1} y \tag{7}$$

$$= K(K + \lambda I)^{-1} y \tag{8}$$

– 5.6.

For a new point $\tilde{x}$

$$\tilde{x}^T w^* = \tilde{x}^T X^T (K + \lambda I)^{-1} y \tag{9}$$

$$= \begin{pmatrix} \tilde{x}^T x_1 & \tilde{x}^T x_2 & \cdots & \tilde{x}^T x_n \end{pmatrix} (K + \lambda I)^{-1} y \tag{10}$$

$$= k_{\tilde{x}}^T (K + \lambda I)^{-1} y \tag{11}$$

6 Decision Trees

• 6.1 Building Trees by Hand.

– 6.1.1.

a) Split on size:
i) Size ≤ 1, $p_1 = \frac{2}{3}, N_1 = 3, Q_1 = \frac{4}{9}, p_2 = \frac{3}{8}, N_2 = 8, Q_2 = \frac{30}{64}, N_1Q_1 + N_2Q_2 = \frac{61}{12} \approx 5.08$
ii) Size ≤ 2, $p_1 = \frac{2}{5}, N_1 = 5, Q_1 = \frac{12}{25}, p_2 = \frac{3}{6}, N_2 = 6, Q_2 = \frac{18}{36}, N_1Q_1 + N_2Q_2 \approx 5.4$
iii) Size ≤ 3, $p_1 = \frac{2}{6}, N_1 = 6, Q_1 = \frac{16}{36}, p_2 = \frac{3}{5}, N_2 = 5, Q_2 = \frac{12}{25}, N_1Q_1 + N_2Q_2 \approx 5.06$
iv) Size ≤ 4, $p_1 = \frac{4}{9}, N_1 = 9, Q_1 = \frac{40}{81}, p_2 = \frac{1}{2}, N_2 = 2, Q_2 = \frac{1}{2}, N_1Q_1 + N_2Q_2 \approx 5.4$

b) split on spots:
v) spots = N, $p_1 = 0, N_1 = 4, Q_1 = 0, p_2 = \frac{5}{7}, N_2 = 7, Q_2 = \frac{20}{49}, N_1Q_1 + N_2Q_2 \approx 2.85$

c) split on color:
vi) color = white, $p_1 = \frac{2}{5}, N_1 = 5, Q_1 = \frac{12}{25}, p_2 = \frac{3}{6}, N_2 = 6, Q_2 = \frac{18}{36}, N_1Q_1 + N_2Q_2 \approx 5.4$
The minimal weighted impurity measure is obtained by splitting on the spots.

$$\begin{tikzcd} & \text{Original} \arrow[dl] \arrow[dr] & \\ \begin{array}{ccc} \multicolumn{3}{c}{\text{Spots} = \text{N}} \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{array} & & \begin{array}{ccc} \multicolumn{3}{c}{\text{Spots} = \text{Y}} \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{array} \end{tikzcd}$$