50.007 Machine Learning
2015 Term 6
Sample midterm solutions

Shaowei Lin
01 Nov 2016

(Q1) b

(Q2) c

we label choices by a, b, c in order of appearance.

(Q3) No

(Q4) $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

(Q5) $\frac{1}{N} \sum_{i=1}^{N} \max\{1 - y_i f(x_i), 0\}$

$\frac{1}{N} \sum_{i=1}^{N} (y_i - f(x_i))^2$

(Q6) (a) $\nabla f(x) = 21x^6 - 25x^4 + 14x$

(b) we write $\sum_{r=1}^{d} x^{(r)} = \mathbb{1}^T x$

where $\mathbb{1}$ is a column vector of ones.

Then,

$$f(x) = x^T (A z \mathbb{1}^T) x$$

where $A z \mathbb{1}^T$ is a square matrix of constants (since we are differentiating with respect to x only).

Note

If A is not symmetric, $\nabla_x (x^T A x) = (A^T + A)x$

Hence, $\nabla f(x) = \cancel{2 A z \mathbb{1}^T x}$ $(A z \mathbb{1}^T + \mathbb{1} z^T A^T) x$

$= \cancel{2 A z \sum_{r=1}^{d} x^{(r)}}$.

(c) we use chain rule.

$$\nabla f(x) = 4\,(x^T A x)^3 \cdot 2Ax$$
$$= 8\,(x^T A x)^3\, Ax$$

(d) we use product rule.

First note that

$$\|Ax\|_2 = (x^T A^T A x)^{1/2}$$

Thus,

$$\nabla f(x) = \tfrac{1}{2}(x^T A^T A x)^{-1/2} \cdot 2A^T A x \cdot x^T x$$
$$+ (x^T A^T A x)^{1/2} \cdot 2x$$
$$= (x^T A^T A x)^{-1/2} \cdot \left[A^T A x \cdot x^T x + 2\, x^T A^T A x \cdot x \right]$$
$$= \frac{1}{\|Ax\|_2} \cdot \left[A^T A x x^T x + 2 x x^T A^T A x \right]$$
$$= \left(A^T A x x^T + 2 x x^T A^T A \right) \frac{x}{\|Ax\|_2}$$

we cannot simplify this to $3 A^T A x x^T$

(Q7) concave

(Q8) None

(Q9) The loss function z^3 is negative for negative errors z. Since the objective is to minimize the loss, the trained parameters will end up favoring large negative errors.

(Q10) $\log p(x) = \log c + 2\log x + 3\log\beta - \beta x$

log likelihood of the data

$$\ell(\beta) = \sum_{i=1}^{n} \log p(x_i)$$

$$= n\log c + 2\sum_{i=1}^{n}\log x_i + 3n\log\beta - \beta\sum_{i=1}^{n} x_i$$

$$\nabla\ell(\beta) = \frac{3n}{\beta} - \sum_{i=1}^{n} x_i$$

solving $\nabla\ell(\hat\beta) = 0$ gives $\hat\beta = \frac{3n}{\sum_{i=1}^{n} x_i}$

To show that this achieves the maximum of $\ell(\beta)$, we compute the Hessian.

$$\nabla^2 \ell(\beta) = -\frac{3n}{\beta^2} < 0.$$

Since the Hessian is negative, $\hat{\beta}$ must be the MLE.

(Q11) (a) Yes, it is a kernel function.

Let $\phi(x) = \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix} x$

Then $k(x_1, x_2) = \langle \phi(x_1), \phi(x_2) \rangle$

$$= x_1^T \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix}^T \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix} x_2$$

$$= x_1^T \begin{pmatrix} 25 & 0 \\ 0 & 9 \end{pmatrix} x_2 .$$

(b) No, it is not a kernel

Let $x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $x_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Then, the Gram matrix for these two points is K where

$$K_{11} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 25 & 0 \\ 0 & -9 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 25$$

$$K_{12} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 25 & 0 \\ 0 & -9 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 0$$

$$K_{21} = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 25 & 0 \\ 0 & -9 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0$$

$$K_{22} = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 25 & 0 \\ 0 & -9 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -9$$

$K = \begin{pmatrix} 25 & 0 \\ 0 & -9 \end{pmatrix}$ is not positive definite, because the eigenvalues are 25 and -9, and -9 is negative.

(Q11) (c) No, it is not a kernel because it is not symmetric.

$$k\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}2 & 0\\1 & 2\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = 0$$

$$k\left(\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}\right) = \begin{pmatrix}0 & 1\end{pmatrix}\begin{pmatrix}2 & 0\\1 & 2\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix} = 1$$
