Exercises
Introduction to Machine Learning
FS 2018

Series 2, Mar 30th, 2018 (Kernel)

Institute for Machine Learning
Dept. of Computer Science, ETH Zürich
Prof. Dr. Andreas Krause
Web: `https://las.inf.ethz.ch/teaching/introml-s18`
Email questions to:
Kjong Lehmann, kjong.lehmann@inf.ethz.ch

Problem 1 (Perceptron/SVM):

Answer 1a:
Perceptron is SGD on the perceptron loss function

$$\nabla l_p(w_i, x, y) = \begin{cases} 0 \text{ if } y = sign(w^T x) \\ -yx \text{ if } y \neq sign(w^T x) \end{cases}$$

Answer 1b:
Perceptron

$$min \sum_i max\{0, -y_i \alpha^T k_i\}$$

SVM

$$min \sum_i max\{0, 1 - y_i \alpha^T k_i\} + \lambda ||w||_2^2$$

Difference is essentially an L2 penalty
Answer 1c:

$$\nabla G(w) = \frac{1}{n} \sum_i^B \nabla g_i(x)$$

$$\nabla g_i(w) = \nabla max(0, 1 - y_i w^T x_i) + \nabla \lambda ||w||_2^2$$

So looking at these separately:

$$\nabla \lambda ||w||^2 = 2\lambda w_k$$

$$\nabla max(0, 1 - y_i w^T x_i) = \begin{cases} 0 \text{ if } y_i w^T x_i \geq 1 \\ -y_i x_i \text{ otherwise} \end{cases}$$

So from this we get:

$$w = w + \eta(-2\lambda w) \text{ if } y_i w^T x_i \geq 1$$
$$w = w + \eta(y_i x_i - 2\lambda w) \text{ otherwise}$$

Answer 2a:
Can show that -f(x) is convex

$$\begin{aligned} \sqrt{tx_1 + (1-t)x_2} &> t\sqrt{x_1} + (1-t)\sqrt{x_2} \\ tx_1 + (1-t)x_2 &> t^2 x_1 + (1-t)^2 x_2 + t2(1-t)\sqrt{x_1 x2} \\ x_1 + x_2 &> 2\sqrt{x_1 x_2} \\ (\sqrt{x_1} - \sqrt{x_2})^2 &> 0 \end{aligned}$$

Answer 2b:
Note that x is restricted to be positive, so one can easily check a few examples of p to gain some intuition.

$$f'' = p(p-1)x^{p-2}$$
$$p(p-1) > 0$$

Since p positiv and x positiv, the second derivated will always be positive

Problem 2 (Feature Selection):

$$\begin{aligned}
J_{EL}(w) &= |y - Xw|^2 + \lambda_2|w|^2 + \lambda_1|w|_1 \\
cJ_{EL}(w) &= c|y - Xw|^2 + c\lambda_2|w|^2 + c\lambda_1|w|_1 \\
&= cy^Ty - 2cy^TXw + cw^TX^TXw + c\lambda_2|w|^2 + c\lambda_1|w|_1 \\
&= cy^Ty - 2cy^TXw + c(w^TX^TXw + \lambda_2|w|^2) + c\lambda_1|w|_1 \\
&= c\tilde{y}^T\tilde{y} - 2\tilde{y}^T\tilde{X}w + c(w^TX^TXw + \lambda_2|w|^2) + c\lambda_1|w|_1 \\
&= c\tilde{y}^T\tilde{y} - 2\tilde{y}^T\tilde{X}w + \frac{1}{c}(w^T\tilde{X}^T\tilde{X}w) + c\lambda_1|w|_1
\end{aligned}$$

Let $\tilde{w} = c^{-1}w$:

$$\begin{aligned}
J_{EL}(\tilde{w}) &= c\tilde{y}^T\tilde{y} - 2c\tilde{y}^T\tilde{X}\tilde{w} + c(\tilde{w}^T\tilde{X}^T\tilde{X}\tilde{w}) + c^2\lambda_1|\tilde{w}|_1 \\
J_{EL}(\tilde{w}) &= c|\tilde{Y} - \tilde{X}w|^2 + c^2\lambda_1|\tilde{w}|_1
\end{aligned}$$

Problem 3 (Kernel):

Answer a:

$$k(\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \\ . \\ . \\ . \\ x'_n \end{pmatrix}) = (1 + x_1x'_1 + x_2x'_2 + ... + x_nx'_n)^2$$

$$\begin{aligned}
&= (1 + (x_1x'_1)^2 + ... + (x_nx'_n)^2 + 2x_1x'_1 + ... + 2x_nx'_n + 2x_1x'_1x_2x'_2 + ... + 2x_1x'_1x_nx'_n + ...) \\
&= \phi(x)^T\phi(x\prime)
\end{aligned}$$

Thus:

$$\phi(x) = (1, \sqrt{2}x_1, ..., \sqrt{2}x_n), \sqrt{2}x_2x_1, \sqrt{2}x_{n-1}x_1, ..., \sqrt{2}x_{n-1}x_{n-2}, \sqrt{2}x_nx_1, , ...\sqrt{2}x_nx_{n-1}, x_1^2, ..., x_n^2)$$

Answer b:

$$\begin{aligned}
k(x, x) &= \phi(x)^T\phi(x') \\
k(x, x) &= \begin{pmatrix} x^{(1)}x^{(1)} + x^{(2)}x^{(2)} + ||x||^2 & x^{(1)}x'^{(1)} + x^{(2)}x'^{(2)} + ||x||||x'|| \\ x^{(1)}x'^{(1)} + x^{(2)}x'^{(2)} + ||x||||x'|| & x'^{(1)}x'^{(1)} + x'^{(2)}x'^{(2)} + ||x\prime||^2 \end{pmatrix} \\
k(x, x) &= \begin{pmatrix} 50 & 2 \\ 2 & 2 \end{pmatrix}
\end{aligned}$$