# Machine Learning
# Homework 2

Due on April 21, 2020

1. Let the sequence P, P, P, P, N, P, P, N, P, N, P, N, P, N, N, N, N, P, N, N be the sorted result according to the *posterior probability* being a positive instance. Please find the AUC value for this ranking result. (10 %)

2. **(a)** Solve

$$\min_{\mathbf{x} \in R^2} \frac{1}{2}\mathbf{x}^\top \begin{bmatrix} 1 & 0 \\ 0 & 1000 \end{bmatrix} \mathbf{x}$$

using the *steep descent with exact line search*. You are welcome to copy the MATLAB code from my slides. Start your code with the initial point $\mathbf{x}^0 = [1000\ 1]^\top$. Stop until $\|\mathbf{x}^{n+1} - \mathbf{x}^n\|_2 < 10^{-8}$. Report your solution and the number of iteration. (10 %)

**(b)** Implement the Newton's method for minimizing a quadratic function $f(x) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + p^\top x$ in MATLAB code. Apply your code to solve the minimization problem in **(a)**.(10 %)

3. Let $S = \{(\mathbf{x}^i, y_i)\}_{i=1}^{\ell} \subseteq R^n \times \{-1, 1\}$ be a non-trivial training set. The Perceptron Algorithm in *dual form* is given as follows:

Given a training set $S$

$\alpha \longleftarrow \mathbf{0}$ and $b \longleftarrow 0$

$L \longleftarrow \max_{1 \le i \le \ell} \|\mathbf{x}^i\|_2$

Repeat

for $i = 1$ *to* $\ell$

if $y_i(\sum_{j=1}^{\ell} \alpha_j y_j \langle \mathbf{x}^i, \mathbf{x}^j \rangle + b) \le 0$

then

$\alpha_i \longleftarrow \alpha_i + 1$

$b \longleftarrow b + y_i L^2$

end if

end for

end for

until no mistakes made within the *for* loop

return $(\alpha, b)$ and define the linear classifier

$$f(x) = \sum_{i=1}^{\ell} \alpha_i y_i \langle \mathbf{x}^i, \mathbf{x} \rangle + b$$

the same x

Suppose that the input training set $S$ is linearly separable.

**(a)** What are the meanings of the output $\alpha_i$ and the 1-norm of $\alpha$? (10%)

**(b)** Why the updating rule is effective? (10%)

4. Let $A_+ = \{(0,0),(0.5,0),(0,0.5),(-0.5,0),(0,-0.5)\}$ and $A_- = \{(0.5,0.5),(0.5,-0.5),(-0.5,0.5),(-0.5,-0.5),(1,0),$ $(0,1),(-1,0),(0,-1)\}$. (50 %)

**(a)** Try to find the hypothesis $h(\mathbf{x})$ by implementing the Perceptron algorithm in the *dual form* and replacing the inner product

$$< x^i, x^j > \text{ by } < x^i, x^j >^2, \text{ and } L = \max_{1 \le i \le \ell} \|x^i\|_2^2$$

**(b)** Generate 10,000 points in the box $[-1.5, 1.5] \times [-1.5, 1.5]$ randomly as a test set. Plug these points into the hypothesis that you got in **(a)** and then plot the points for which $h(x) > 0$ with $+$.

**(c)** Repeat **(a)** and **(b)** by using the training data

$$B_+ = \{(0.5,0),(0,0.5),(-0.5,0),(0,-0.5)\} \text{ and}$$

$$B_- = \{(0.5,0.5),(0.5,-0.5),(-0.5,0.5),(-0.5,-0.5)\}.$$

**(d)** Let the nonlinear mapping $\phi : R^2 \to R^4$ defined by

$$\phi(\mathbf{x}) = [-x_1x_2,\ x_1^2,\ x_1x_2,\ x_2^2\ ]$$

Map the training data $A_+$ and $A_-$ into the feature space using this nonlinear map. Find the hypothesis $f(x)$ by implementing the Perceptron algorithm in the *primal form* in the feature space.

**(e)** Repeat **(b)** by using the hypothesis that you got in **(d)**. Please know that you need to map the points randomly generated in **(b)** by the nonlinear mapping $\phi$ first.