Machine Learning
Homework 2

Due on April 12, 2017

1. Let $f : R^n \to R$ be a *strictly convex* function *i.e.*,

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y).$$

   If x^* is a local minimizer then x^* is the unique global minimizer.

2. **(a)** Solve

$$\min_{x \in R^2} \; \frac{1}{2} x^\top \begin{bmatrix} 1 & 0 \\ 0 & 900 \end{bmatrix} x$$

   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 $x^0 = [1000\ 1]^\top$. Stop until $\|x^{n+1} - x^n\|_2 < 10^{-8}$. Report your solution and the number of iteration.

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

3. Let

$$A_+ = \{(0,0), (0.5,0), (0,0.5), (-0.5,0), (0,-0.5)\}$$

and

$$A_- = \{(0.5,0.5), (0.5,-0.5), (-0.5,0.5), (-0.5,-0.5), (1,0), (0,1), (-1,0), (0,-1)\}.$$

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

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

(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 hyposthesis 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 ϕ first.

4. The following table gives 10-fold cross-validation testing accuracy results:

fold	1	2	3	4	5	6	7	8	9	10
Method A	84%	88%	76%	86%	85%	90%	72%	87%	77%	82%
Method B	72%	84%	82%	80%	81%	80%	75%	86%	75%	78%

(a) Can you conclude that the Method A is *different from* Method B with 95% confidence level?

(b) Can you conclude that the Method A is *different from* Method B with 90% confidence level?

5. Given four points in R^3 as follows,

$$S = \{(1,1,1),\ (1,1,-1),\ (-1,-1,1),\ (-1,-1,-1)\} \in R^3.$$

Will these four points be *shattered* by a hyperplane in R^3 real space? Can you make the conclusion that the VC-dimension for hyperplanes in R^3 is strictly less than 4 ? Justify your answer.