Machine Learning
Homework 2

Due on May 3, 2018

1. 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%	70%	84%	75%	78%

   (a) Can you conclude that the Method A is better than Method B with 95% confidence level?

   (b) Can you conclude that the Method A is better than Method B with 90% confidence level?

2. Let the sequence P, P, P, P, N, P, P, P, N, 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.

3. Let $S = \{(\mathbf{x}^i, y_i)\}_{i=1}^m \subseteq R^n \times \{-1, 1\}$ 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 $b \longleftarrow 0$

$L \longleftarrow \max\limits_{1 \le i \le m} \|\mathbf{x}^i\|_2$

Repeat

for $i = 1$ *to* m

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

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 (α, b) and define the linear classifier

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

Suppose that the input training set S is linearly separable.

(a) What are the meanings of the output α_i and the 1-norm of α? (13%)

(b) Why the updating rule is effective? (12%)

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

(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.