CS246 Homework 4 Answers

Charlie Zhang

Feb 2013

1 Question 1 –Support Vector Machine

1.1 (a)

The example:

(0, 0): -1

(0, 1): -1

(1, 0): 1

(1, 1): 1

(2, 0): -1

This is unfeasible under hard constraints SVM but feasible under soft margin SVM.

Such as: $w = (2, 0), b = -1, \xi_1, .., \xi_5 = (0, 0, 0, 0, 4)$

1.2 (b)

Lets set:

$w_j = 0, \forall j = 1, ..., d,$

$b = 0,$

$\xi_i = 1, \forall i = 1, ..., n$

Then $y_i(w \cdot x + b) \geq 1 - \xi_i$ holds true for all i.

1.3 (c)

Let T be the training set and E be the set of points where linear classification mis-classified.

Then $y_i(x_i \cdot w + b) < 0, \forall i \in E$.

Also, $(w, \xi_i, ..., \xi_n)$ is a feasible point, so $y_i(x_i \cdot w + b) \geq 1 - \xi_i, \forall i \in T$ Now we have $1 - \xi_i < 0$, i.e., $\xi_i > 1$, $\forall i \in E$

So $\sum\limits_{i \in T} \xi_i >= \sum\limits_{i \in E} \xi_i > |E|$

1.4 (d)

$\nabla_b f(w, b) = \frac{\delta f(w,b)}{\delta b} = C \sum\limits_{i=1}^{n} \frac{\delta L(x_i, y_i)}{\delta b}$

Where $\frac{\delta L(x_i, y_i)}{\delta b} = \begin{cases} 0 & \text{if } y_i(w \cdot x_i + b) \geq 1 \\ -y_i & \text{otherwise} \end{cases}$

1.5 (e)

2 Question 2 –Decision Tree Learning

2.1 (a)

$G = max[I(D) - (I(D_L) + I(D_R))]$

There we only consider one attribute, so:

$G = I(D) - (I(D_L) + I(D_R)) =$

$|D| \times (1 - \sum\limits_i p_i^2) - |D_L| \times (1 - \sum\limits_i p_{L(i)}^2) - |D_R| \times (1 - \sum\limits_i p_{R(i)}^2) =$

$|x+y+u+v| * ((1-\frac{(x+u)^2}{(x+y+u+v)^2}) + (1-\frac{(y+v)^2}{(x+y+u+v)^2})) - |x+y| * ((1-\frac{x^2}{x+y)^2}) +$

$(1-\frac{y^2}{x+y)^2})) - |u+v| * ((1-\frac{u^2}{u+v)^2}) + (1-\frac{v^2}{u+v)^2})) =$

$\frac{x^2+y^2}{x+y} + \frac{u^2+v^2}{u+v} - \frac{(x+u)^2+(y+v)^2}{x+y+u+v} > 0$

Solve the inequality and we get $(xv-yu)^2 > 0$.

So $\frac{x}{y} \neq \frac{u}{v}$.

2.2 (b)

based on the equation in (a), we have:

$G_{wine} = \frac{30^2+20^2}{50} + \frac{30^2+20^2}{50} - \frac{60^2+40^2}{100} = 0$

$G_{running} = \frac{20^2+10^2}{30} + \frac{40^2+30^2}{70} - \frac{60^2+40^2}{100} = 0.381$

$G_{pizza} = \frac{50^2+30^2}{80} + \frac{10^2+10^2}{20} - \frac{60^2+40^2}{100} = 0.500$

We should use the attribute 'likes pizza'

3 Question 3 –Clustering Data Streams

3.1 (a)

TODO

3.2 (b)

TODO

$2 \cdot cost_w(\hat{S}, T) + 2\sum_{i=1}^{l} cost(S_i, T_i) =$

$2\sum_{i=1}^{l}\sum_{j=1}^{k} |S_{ij}| d(t_{ij}, T)^2 + 2\sum_{i=1}^{l}\sum_{j=1}^{k}\sum_{x \in S_{ij}} d(x, t_{ij})^2 =$

$\sum_{i=1}^{l}\sum_{j=1}^{k}\sum_{x \in S_{ij}} (2d(x, t_{ij})^2 + 2d(t_{ij}, T)^2) \geq$

$\sum_{i=1}^{l}\sum_{j=1}^{k}\sum_{x \in S_{ij}} d(x, T)^2 =$

$cost(S, T)$

3.3 (c)

The code:

```
    print S_bar
  print list_rho
  print list_num_edge
  print list_size_s

if __name__ == '__main__':
  main()
```