Machine Learning, 2021 Spring
Homework 5

Due on 12:59 MAY 17, 2021

Problem 1

Definition 1 (leave-one-out cross-validation) Select each training example in turn as the single example to be held-out, train the classifier on the basis of all the remaining training examples, test the resulting classifier on the held-out example, and count the errors.

Let the superscript '$-i$' denote the parameters we would obtain by finding the SVM classifier f without the ith training example. Define the leave-one-out CV error as

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(y_i, f(\boldsymbol{x}_i; \boldsymbol{w}^{-i}, b^{-i})),$$

where $\mathcal{L}$ is the zero-one loss. Prove that [1.5pts]

$$\text{leave-one-out CV error} \leq \frac{\text{number of support vectors}}{\text{n}}. \tag{1}$$

1. SVM: $\frac{1}{2}||w||_2^2$

s.t. $y_i(w^T x_i + b) \geq 1 \quad i=1,\cdots n$

we let the optimal solution $(w,b)=(w^*,b^*)$ and the region for w can find hyperplane to classifer correctly: $[w^{r_1}, w^{r_2}]$

And define SV set: A (equals to could find b that satisfy all constraints)

not SV set: $B = X \setminus A$, $|A|+|B|=|X|=n$

$y_{i_a}(w^T x_{i_a}+b)=1 \quad \forall x_{i_a} \in A$

$y_{i_b}(w^T x_{i_b}+b) > 1 \quad \forall x_{i_b} \in B$

Analysis:

For leave-one-out cross validation:

If $i \in B$, then w^* satisfy $y_i(w^T x_i + b) \geq 1 \; \forall i=1,\cdots i_b-1, i_b+1,\cdots n$

and since $i \notin B$, the classifer correctly region for w: $[w^{r_1}, w^{r_2}]$ doesn't change.

$\therefore (w^{-i})^* = w^*$

$\therefore \mathcal{L}(y_{i_b}, f(x_i, w^{-i_b}, b^{-i_b})) = 0, \; \forall i_b \in B$

$\therefore \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(y_i, f(x_i, w^{-i}, b^{-i})) = \frac{1}{n}\sum_{\substack{i=1 \\ i \notin B}}^{n}\mathcal{L}(y_i, f(x_i, w^{-i}, b^{-i})) \leq \frac{|A|}{n} \rightarrow$ number of support vectors

Problem 2

The ℓ_1-norm SVM can be formulated as follows

$$\begin{aligned} \min_{(\boldsymbol{w},b)} \quad & \|\boldsymbol{w}\|_1 \\ \text{s.t.} \quad & y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b) \geq 1, \quad i = 1, \cdots, n. \end{aligned} \tag{2}$$

Please derive the equivalent linear programming formulation of (2). [1.5pts]

2. $\min_{(w,b)} \|w\|_1$ s.t. $y_i(w^T x_i + b) \geq 1, i=1,\cdots n$ $\Leftrightarrow$ $\min_{(w,b)} \sum_{i=1}^{n} |w_i|$ s.t. $y_i(w^T x_i + b) \geq 1, i=1,\cdots n$ $\overset{\boxed{LP}}{\Leftrightarrow}$ $\min_{t} \sum_{i=1}^{n} t_i$

$$\begin{aligned} \text{s.t.} \quad & y_i(w^T x_i + b) \geq 1, i=1,\cdots n \\ & w_i \leq t_i \quad i=1,\cdots n \\ & w_i \geq -t_i \quad i=1,\cdots n \end{aligned}$$

Problem 3

For the example in page 14 of Lecture 13, given

$$\boldsymbol{x} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \\ 3 & 0 \end{bmatrix} \quad \boldsymbol{y} = \begin{bmatrix} -1 \\ -1 \\ +1 \\ +1 \end{bmatrix},$$

please provide the soft-margin SVM model of this problem. Derive the associated Lagrangian and the dual problem of it. [3pts]

(Hint: the dual problem is a quadratic programming problem.)

3. $\boxed{\text{Soft margin SVM}}$ $\min\limits_{(w,\xi,b)} \frac{1}{2}\|w\|_2^2 + C\sum\limits_{i=1}^{4}\xi^i$, $C>0$ is a constant

s.t. $-b \geq 1-\xi^1$

$-(2w_1+2w_2+b) \geq 1-\xi^2$

$2w_1+b \geq 1-\xi^3$

$3w_1+b \geq 1-\xi^4$

$\xi^i \geq 0 \quad i=1,\cdots 4$

$$L(w,\xi,\mu) = \frac{1}{2}(w_1^2+w_2^2)+C(\xi^1+\xi^2+\xi^3+\xi^4)+\mu_1(1-\xi^1+b)+\mu_2(1-\xi^2+2w_1+2w_2+b)+\mu_3(1-\xi^3-2w_1-b)+\mu_4(1-\xi^4-3w_1-b)$$

$$-\mu_5\xi^1-\mu_6\xi^2-\mu_7\xi^3-\mu_8\xi^4$$

$$=\frac{1}{2}(w_1^2+w_2^2)+(C-\mu_1-\mu_5)\xi^1+(C-\mu_2-\mu_6)\xi^2+(C-\mu_3-\mu_7)\xi^3+(C-\mu_4-\mu_8)\xi^4$$

$$+(2\mu_2-2\mu_3-3\mu_4)w_1+2\mu_2w_2+(\mu_1+\mu_2-\mu_3-\mu_4)b+\mu_1+\mu_2+\mu_3+\mu_4$$

$$g(\mu)=\inf_{(w,\xi,b)\in D} L(w,\xi,\mu) \Rightarrow \inf_{(w,\xi,b)\in D} \frac{1}{2}(w_1^2+2(2\mu_2-2\mu_3-3\mu_4)+(2\mu_2-2\mu_3-3\mu_4)^2)+\frac{1}{2}(w_2^2+4\mu_2w_2+4\mu_2^2)+(\mu_1+\mu_2-\mu_3-\mu_4)b$$

$$+(C-\mu_1-\mu_5)\xi^1+(C-\mu_2-\mu_6)\xi^2+(C-\mu_3-\mu_7)\xi^3+(C-\mu_4-\mu_8)\xi^4$$

$$-\frac{1}{2}(2\mu_2-2\mu_3-3\mu_4)^2-2\mu_2^2+\mu_1+\mu_2+\mu_3+\mu_4$$

$\therefore$ Dual problem: $\max\limits_{\mu} -\frac{1}{2}(2\mu_2-2\mu_3-3\mu_4)^2-2\mu_2^2+\mu_1+\mu_2+\mu_3+\mu_4$

s.t. $\mu_1+\mu_5=C$

$\mu_2+\mu_6=C$

$\mu_3+\mu_7=C$

$\mu_4+\mu_8=C$

$\mu_1+\mu_2-\mu_3-\mu_4=0$

$\mu_1,\cdots \mu_8 \geq 0$

Problem 4

Complete the decision trees on the following example by both ID3 and CART methods (refer to Lecture 14 for more details). [4pts]

outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

4. ID3 $E = \sum_x p(x) \log_2 \frac{1}{p(x)}$

1° $E(S) = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.9403$

1' $E(S, \text{outlook} = \text{"sunny"}) = -\frac{2}{5}\log_2(\frac{2}{5}) - \frac{3}{5}\log_2(\frac{3}{5}) = 0.971$

$E(S, \text{outlook} = \text{"overcast"}) = 0$

$E(S, \text{outlook} = \text{"rainy"}) = 0.971$

$\text{Gain}(\text{outlook}) = E(S) - (\frac{5}{14}\times 0.971 + \frac{4}{14}\times 0 + \frac{5}{14}\times 0.971) = 0.247$

2' $E(S, \text{temperature} = \text{"hot"}) = -\frac{2}{4}\log_2\frac{2}{4} \times 2 = 1$

$E(S, \text{temperature} = \text{"mild"}) = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} = 0.918$

$E(S, \text{temperature} = \text{"cool"}) = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4} = 0.811$

$\text{Gain}(\text{temperature}) = E(S) - (\frac{4}{14}\times 1 + \frac{6}{14}\times 0.918 + \frac{4}{14}\times 0.811) = 0.029$

3' $E(S, \text{humidity} = \text{"high"}) = -\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7} = 0.985$

$E(S, \text{humidity} = \text{"normal"}) = -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} = 0.592$

$\text{Gain}(\text{humidity}) = E(S) - (\frac{1}{2}\times 0.985 + \frac{1}{2}\times 0.592) = 0.1518$

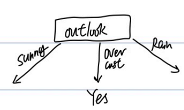
4' $E(S, \text{windy} = \text{"FALSE"}) = -\frac{6}{8}\log_2\frac{6}{8} - \frac{2}{8}\log_2\frac{2}{8} = 0.811$

$E(S, \text{windy} = \text{"TRUE"}) = -\frac{3}{6}\log_2\frac{3}{6} \times 2 = 1$

$\text{Gain}(\text{windy}) = E(S) - (\frac{8}{14}\times 0.811 + \frac{6}{14}\times 1) = 0.0483$

$\text{Gain}^*(x) = \text{Gain}(\text{outlook}) = 0.247$

∴



2° for sunny, $E(S) = 0.971$

1' $E(S, \text{temperature} = \text{"hot"}) = 0$

$E(S, \text{temperature} = \text{"mild"}) = 1$

$E(S, \text{temperature} = \text{"cool"}) = 0$

$\text{Gain}(\text{temperature}) = E(S) - 1\times\frac{2}{5} = 0.571$

2' $E(S, \text{humidity} = \text{"high"}) = 0$

$E(S, \text{humidity} = \text{"normal"}) = 0$

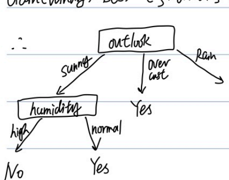
$\text{Gain}(\text{humidity}) = E(S) - 0 = 0.971$

3' $E(S, \text{windy} = \text{"FALSE"}) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.918$

$E(S, \text{windy} = \text{"TRUE"}) = 1$

$\text{Gain}(\text{windy}) = E(S) - (\frac{3}{5}\times 0.918 + \frac{2}{5}\times 1) = 0.0202$ $\text{Gain}^*(x) = \text{Gain}(\text{humidity}) = 0.971$

∴



for rainy, $E(S)=0.971$

1' $E(S, temperature="mild") = -\frac{1}{3}\log_2\frac{2}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.918$

$E(S, temperature="cool") = 1$

$Gain(temperature) = E(S) - (\frac{3}{5}\times 0.918 + \frac{2}{5}\times 1) = 0.0202$

2' $E(S, humidity="high") = 1$

$E(S, humidity="normal") = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.918$

$Gain(humidity) = E(S) - (\frac{2}{5}\times 1 + \frac{3}{5}\times 0.918) = 0.0202$

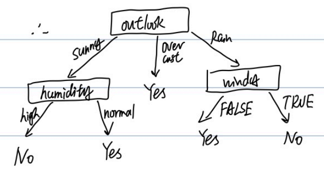
3' $E(S, windy="FALSE") = 0$

$E(S, windy="TRUE") = 0$

$Gain(windy) = E(S) - 0 = 0.971$

$Gain^*(x) = Gain(windy) = 0.971$

$\therefore$



CART $E = \sum_x p(x)(1-p(x))$

1° $E(S) = \frac{9}{14}\times\frac{5}{14} + \frac{5}{14}\times\frac{9}{14} = 0.459$

1' $E(S, outlook="sunny") = \frac{2}{5}\times\frac{3}{5}\times 2 = 0.48$

$E(S, outlook="overcast") = 0$

$E(S, outlook="rainy") = \frac{2}{5}\times\frac{3}{5}\times 2 = 0.48$

$Gini(outlook) = E(S) - (\frac{5}{14}\times 0.48 + \frac{4}{14}\times 0 + \frac{5}{14}\times 0.48) = 0.116$

2' $E(S, temperature="hot") = \frac{1}{2}\times\frac{1}{2}\times 2 = \frac{1}{2}$

$E(S, temperature="mild") = \frac{4}{6}\times\frac{2}{6}\times 2 = \frac{4}{9}$

$E(S, temperature="cool") = \frac{3}{4}\times\frac{1}{4}\times 2 = \frac{3}{8}$

$Gini(temperature) = E(S) - (\frac{4}{14}\times\frac{1}{2} + \frac{6}{14}\times\frac{4}{9} + \frac{4}{14}\times\frac{3}{8}) = 0.0185$

3' $E(S, humidity="high") = \frac{3}{7}\times\frac{4}{7}\times 2 = \frac{24}{49}$

$E(S, humidity="normal") = \frac{6}{7}\times\frac{1}{7}\times 2 = \frac{12}{49}$

$Gini(humidity) = E(S) - (\frac{1}{2}\times\frac{24}{49} + \frac{1}{2}\times\frac{12}{49}) = 0.0917$

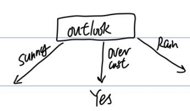
4' $E(S, windy="FALSE") = \frac{6}{8}\times\frac{2}{8}\times 2 = \frac{3}{8}$

$E(S, windy="TRUE") = \frac{1}{2}\times\frac{1}{2}\times 2 = \frac{1}{2}$

$Gini(windy) = E(S) - (\frac{8}{14}\times\frac{3}{8} + \frac{6}{14}\times\frac{1}{2}) = 0.0304$

$Gini^*(x) = Gini(outlook) = 0.116$

$\therefore$



2° for sunny, $E(S) = 0.48$

1' $E(S, temperature="hot") = 0$

$E(S, temperature="mild") = \frac{1}{2}\times\frac{1}{2}\times 2 = \frac{1}{2}$

2' $E(S, humidity="high") = 0$

$E(S, humidity="normal") = 0$

$E(S, temperature=\text{"cool"}) = 0$

$Gini(humidity) = E(S) - 0 = 0.48$

$Gini(temperature) = E(S) - \frac{2}{5} \times \frac{1}{2} = 0.28$

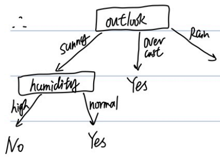
3' $E(S, windy=\text{"FALSE"}) = \frac{1}{3} \times \frac{2}{3} \times 2 = \frac{4}{9}$

$E(S, windy=\text{"TRUE"}) = \frac{1}{2} \times \frac{1}{2} \times 2 = \frac{1}{2}$

$Gini(windy) = E(S) - (\frac{3}{5} \times \frac{4}{9} + \frac{2}{5} \times \frac{1}{2}) = 0.0133$

$Gini^*(x) = Gini(humidity) = 0.48$

$\therefore$



for rainy, $E(S) = 0.48$

1' $E(S, temperature=\text{"mild"}) = \frac{1}{3} \times \frac{2}{3} \times 2 = \frac{4}{9}$

$E(S, temperature=\text{"cool"}) = \frac{1}{2} \times \frac{1}{2} \times 2 = \frac{1}{2}$

$Gini(temperature) = E(S) - (\frac{3}{5} \times \frac{4}{9} + \frac{2}{5} \times \frac{1}{2}) = 0.0133$

2' $E(S, humidity=\text{"high"}) = \frac{1}{2} \times \frac{1}{2} \times 2 = \frac{1}{2}$

$E(S, humidity=\text{"normal"}) = \frac{1}{3} \times \frac{2}{3} \times 2 = \frac{4}{9}$

$Gini(humidity) = E(S) - (\frac{2}{5} \times \frac{1}{2} + \frac{3}{5} \times \frac{4}{9}) = 0.0133$

3' $E(S, windy=\text{"FALSE"}) = 0$

$E(S, windy=\text{"TRUE"}) = 0$

$Gini(windy) = E(S) - 0 = 0.48$

$Gini^*(x) = Gini(windy) = 0.48$

$\therefore$