CS578/STAT590: Introduction to Machine Learning **Fall 2014**

Problem Set 2

Gen Nishida *Handed In: October 1, 2014*

1 Questions

1. (1) Boolean function

   By using conjunctions for every combination of three variables, we can test if at least three variables are active. Therefore, the Boolean function for this is as follows:

$$f = (x_1 \wedge x_2 \wedge x_3) \vee (x_1 \wedge x_2 \wedge x_4) \vee (x_1 \wedge x_2 \wedge x_5) \vee (x_1 \wedge x_3 \wedge x_4) \vee (x_1 \wedge x_3 \wedge x_5)$$
$$\vee (x_1 \wedge x_4 \wedge x_5) \vee (x_2 \wedge x_3 \wedge x_4) \vee (x_2 \wedge x_3 \wedge x_5) \vee (x_2 \wedge x_4 \wedge x_5) \vee (x_3 \wedge x_4 \wedge x_5)$$

   (2) Linear function

   Linear function that can achieve the same classification just needs to checke if the sum is more than two.

$$f = \begin{cases} 1 & (x_1 + x_2 + x_3 + x_4 + x_5 \geq 3) \\ 0 & Otherwise \end{cases}$$

2. What is the size of CON_B?

   For every variable, there are two cases, used in the conjunctions or not used. Thus, the size of CON_B is 2^n.

3. What is the size of CON_L?

   For every pair of $f_b^{(i)}$ and $f_b^{(j)}$ $(i \neq j)$ in CON_B, $\exists x, f_b^{(i)} \neq f_b^{(j)}$. Thus, ths size of CON_L has to be larger than the size of CON_B to make it consistent with CON_B. However,

4. Mistake bound

   Mistake bound is the maximum possible number of mistakes made by the online learning algorithms, which is also used to evaluate the performance of the convergence of the algorithms.

5. mistake bound algorithm

   1) initialize the hypothesis: $f = x_1 \wedge \neg x_1 \wedge x_2 \wedge \neg x_2 \wedge \cdots \wedge x_n \wedge \neg x_n$
   2) if no mistake do nothing
   3) else
   4) for all in-active variables, remove x_i from the conjunctions

   The example of the learning process is shown in Figure 1. For instance, when the first mistake is made for an example $x = (1, 0, 0)$, the variables $\neg x_1$, x_2, and x_3 are in-active, and these variables are removed from the conjunctions, which results in the updated hypothesis $f = x_1 \wedge \neg x_2 \wedge \neg x_3$.

For every mistake, we remove at least one unnecessary variable from the conjunctions. Since we have at least one variable in the conjunctions, the total number of mistakes is at most $2n - 1$. Thus, this is a mistake bound algorithm.

$f = x_1 \wedge \neg x_2$

Initialize: $\hat{f} \leftarrow x_1 \wedge \neg x_1 \wedge x_2 \wedge \neg x_2 \wedge x_3 \wedge \neg x_3$

$\langle(1,1,1), -\rangle$ $\hat{f}(1,1,1) = 0$ ok

$\langle(1,0,0), +\rangle$ $\hat{f}(1,0,0) = 0$ mistake

In-active variables $\neg x_1, x_2, x_3$ are removed:

$\hat{f} \leftarrow x_1 \wedge \neg x_2 \wedge \neg x_3$

$\langle(1,1,1), -\rangle$ $\hat{f}(1,1,1) = 0$ ok

$\langle(1,0,1), +\rangle$ $\hat{f}(1,0,1) = 0$ mistake

In-active variable $\neg x_3$ is removed:

$\hat{f} \leftarrow x_1 \wedge \neg x_2$

Figure 1: The example of a learning process

6. (1) Will both classifiers converge?

   Yes. For the linearly separable dataset, it is proved that the perceptron algorithm will converge, and the mistake bound is R^2/γ^2, which is independent from the order of the examples.

   (2) What will be the training error of each one of the classifiers?

   When both classifiers converge, no more mistakes will be made. In other words, the training error will be zero for both classifiers.

7. kernel function $K(x, y)$

   x and y have the $same(x, y)$ values in common in their variables. Since we have at most k variables in each conjunction, the number of all the combination of the same variables of size at most k will be the kernel function. Thus,

$$K(x, y) = {}_{same(x,y)}C_1 + {}_{same(x,y)}C_2 + \cdots + {}_{same(x,y)}C_k$$

2 Programming Assignment

1. Feature Representation

2. Implementation

3. Experiments