Ben Holden

Robotics 831, Homework #3

11/09/2015

1.) List Prediction
1.2) Theory Questions
1.2.1) Monotone Submodularity
Multiple Guess:

$$f(L;T) = \min(1, |L \cap T|)$$

T is a subset of S
T is the set of all good candidates
L is the set of selected candidates
If L has no overlap with T then no good candidates have been selected and f returns 0, else 1
Monotone:

$$f(L_1) \leq f(L_1 \oplus L_2)$$
$$f(L_2) \leq f(L_2 \oplus L_1)$$

4 Cases:
Let L_1 be in set T be shown as $L_1 = 1$, else $L_1 = 0$
Let L_2 be in set T be shown as $L_2 = 1$, else $L_2 = 0$
Let $f(L_1) \leq f(L_1 \oplus L_2)$ be A
Let $f(L_2) \leq f(L_2 \oplus L_1)$ be B

L_1	L_2	$f(L_1)$	$f(L_2)$	$f(L_1 \oplus L_2)$	$f(L_2 \oplus L_1)$	A	B
0	0	0	0	0	0	1	1
0	1	0	1	1	1	1	1
1	0	1	0	1	1	1	1
1	1	1	1	1	1	1	1

A and B are always 1, proving it is Monotone.
Submodularity:

$$b(s|L_1) \geq b(s|L_1 \oplus L_2)$$

Where

$$b(s|L) = f(L \oplus s) - f(L)$$

Prove:

$$f(L_1 \oplus s) - f(L_1) \geq f(L_1 \oplus L_2 \oplus s) - f(L_1 \oplus L_2)$$

Eight Cases:
Let s be in set T be shown as $s = 1$, else $s = 0$
Let L_1 be in set T be shown as $L_1 = 1$, else $L_1 = 0$
Let L_2 be in set T be shown as $L_2 = 1$, else $L_2 = 0$
Let $f(L_1 \oplus s) - f(L_1)$ be A
Let $f(L_1 \oplus L_2 \oplus s) - f(L_1 \oplus L_2)$ be B
Let $f(L_1 \oplus s) - f(L_1) \geq f(L_1 \oplus L_2 \oplus s) - f(L_1 \oplus L_2)$ be C

s	L_1	L_2	$f(L_1)$	$f(L_1 \oplus s)$	$f(L_1 \oplus L_2)$	D$f(L_1 \oplus L_2 \oplus s)$	A	B	C
0	0	0	0	0	0	0	0	0	1
0	0	1	0	0	1	1	0	0	1
0	1	0	1	1	1	1	0	0	1
0	1	1	1	1	1	1	0	0	1
1	0	0	0	1	0	1	1	1	1
1	0	1	0	1	1	1	1	0	1
1	1	0	1	1	1	1	0	0	1
1	1	1	1	1	1	1	0	0	1

G is always 1 so:

$$f(L_1 \oplus s) - f(L_1) \geq f(L_1 \oplus L_2 \oplus s) - f(L_1 \oplus L_2)$$

is always true, proving the submodularity
1.2.2) Greedy Guarantee

$$L_{i+1}^G = L_i^G \oplus \arg\max_{s \epsilon S} b\left(s|L_i^G\right)$$

Step 1: Prove:

$$\triangle_i \leq \sum_{j=1}^{k} \left\{ f\left(L_{i-1}^G \oplus l_j^*\right) - f\left(L_{i-1}^G\right) \right\}$$

Where:

$$\triangle_i = f(L^*) - f(L_{i-1}^G)$$

Proof:

$$f(L^* \oplus L_{i-1}^G) = f(L^* \oplus L_{i-1}^G)$$

$$f(L^* \oplus L_{i-1}^G) = f(L^* \oplus L_{i-1}^G) + f(L_{i-1}^G \oplus \{l_1^*, ..., l_{k-1}^*\}) - f(L_{i-1}^G \oplus \{l_1^*, ..., l_{k-1}^*\})$$

$$f(L^* \oplus L_{i-1}^G) - f(l_{i-1}^G) = \sum_{j=1}^{k} \left(f(L_{i-1}^G \oplus \{l_1^*, ..., l_j^*\}) - f(L_{i-1}^G \oplus \{l_1^*, ..., l_{j-1}^*\}) \right)$$

By submodularity:

$$f(L^* \oplus L_{i-1}^G) - f(l_{i-1}^G) = \sum_{j=1}^{k} \left\{ f\left(L_{i-1}^G \oplus l_j^*\right) - f\left(L_{i-1}^G\right) \right\}$$

By monotonicity:

$$f(L^*) \leq f(L^* \oplus L_{i-1}^G)$$

So:

$$f(L^*) - f(l_{i-1}^G) \leq \sum_{j=1}^{k} \left\{ f\left(L_{i-1}^G \oplus l_j^*\right) - f\left(L_{i-1}^G\right) \right\}$$

By definition of $\triangle_i$:

$$\triangle_i \leq \sum_{j=1}^{k} \left\{ f\left(L_{i-1}^G \oplus l_j^*\right) - f\left(L_{i-1}^G\right) \right\}$$

Step 2: Prove:

$$\triangle_{i+1} \leq (1 - 1/k)\, \triangle_i$$

Proof:

$$f(L^*) - f(l_{i-1}^G) \leq \sum_{j=1}^{k} \left\{ f\left(L_{i-1}^G \oplus l_j^*\right) - f\left(L_{i-1}^G\right) \right\}$$

Chosen greedily so at most:

$$\sum_{j=1}^{k} \left\{ f\left(L_{i-1}^G \oplus l_j^*\right) - f\left(L_{i-1}^G\right) \right\} = k\left(f(L_i^G) - f(L_{i-1}^G)\right)$$

So:

$$f(L^*) - f(l_{i-1}^G) \leq k\left(f(L_i^G) - f(L_{i-1}^G)\right)$$

Rearranged:

$$f(L^*) - f(l_i^G) \leq (1 - 1/k)\left(f(L^*) - f(l_{i-1}^G)\right)$$

Substitute for $\triangle_i$ and $\triangle_{i+1}$:

$$\triangle_{i+1} \leq (1 - 1/k)\,\triangle_i$$

Step 3: Prove:

$$f(L^G) \geq (1 - 1/e) f(L^*)$$

Where:

$$\triangle_1 = f(L^*)$$

$$\triangle_{k+1} = f(L^*) - f(L^G)$$

$$(1 - 1/n)^n \leq 1/e$$

Proof:

$$\triangle_{i+1} \leq (1 - 1/k)\,\triangle_i$$

Using induction:

$$\triangle_{i+1} \leq (1 - 1/k)^i\,\triangle_1$$

When $i = k$:

$$\triangle_{k+1} \leq (1/e)\,\triangle_1$$

Plug in for $\triangle_1$ and $\triangle_{k+1}$:

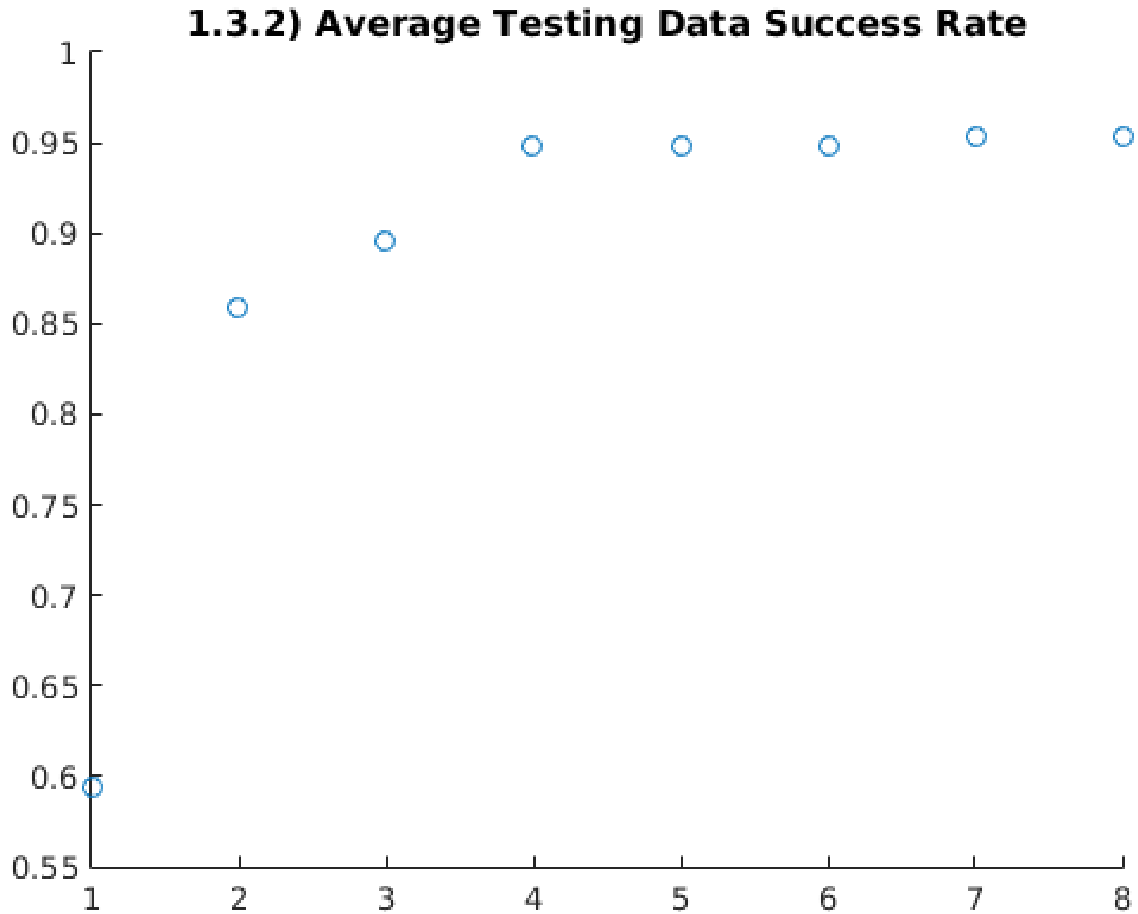
$$f(L^*) - f(L^G) \leq (1 - 1/k)\,f(L^*)$$

Rearrage/algebra:

$$\boxed{f(L^G) \geq (1 - 1/e) f(L^*)}$$

1.3) Coding
1.3.2) Naive Prediction Strategy

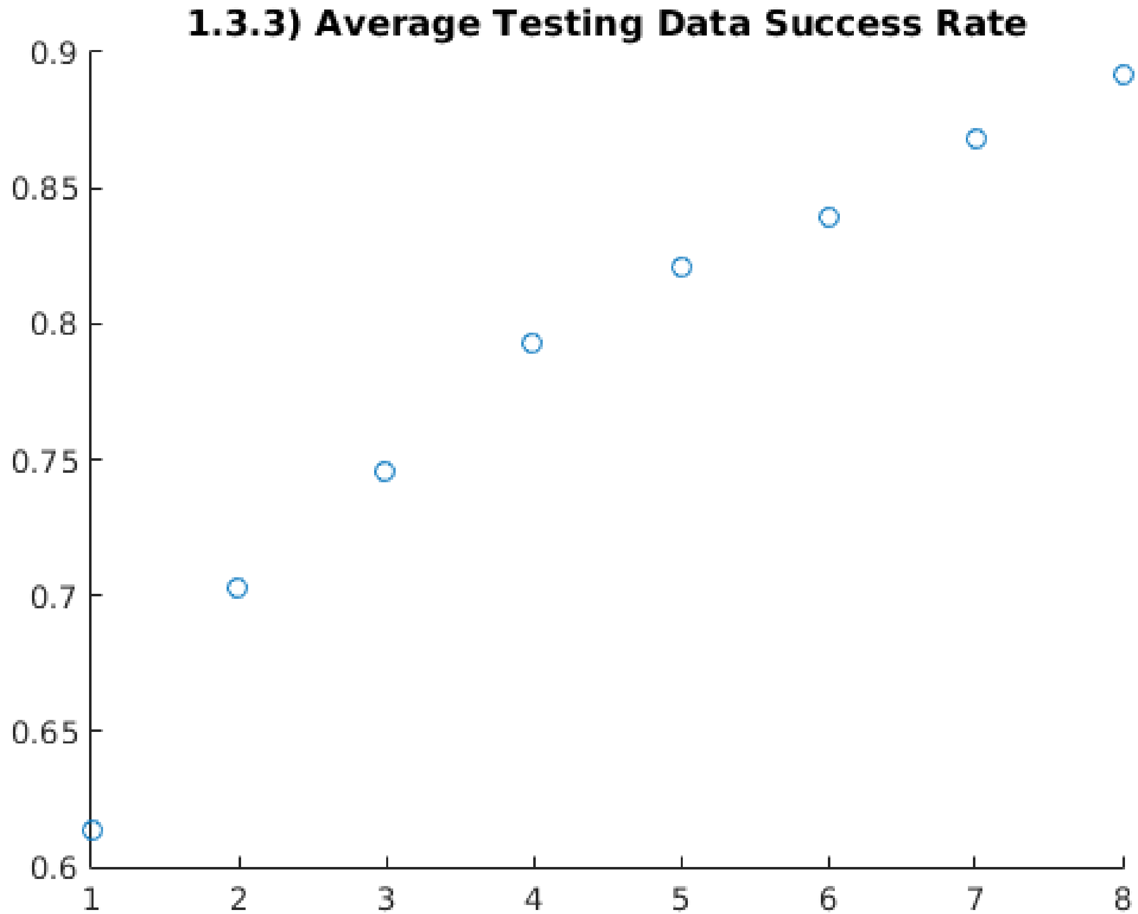

1.3.2) Average Training Data Success Rate
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
1
2
3
4
5
6
7
8



Using SVM to learn the data resulted with a 33.30% misclassification error. The learner does ok with just one element and as the list grows, the success rate goes up peaking at around 95%.

1.3.3) List Prediction Strategy


1.3.3) Average Training Data Success Rate
0.95
0.9
0.85
0.8
0.75
0.7
0.65
1
2
3
4
5
6
7
8



These results show that by adding diversity to the choices this gives a lot of benefit for each item added to the list. So as k grows in size the average success rate increases by about the same amount as the last time an item was added. This will eventually fall off with enough items because the utility of adding another item will eventually be 0 or close to 0.

2.) Online Support Vector Machines

2.2) Derivation

2.2.1) Prove EQ2 and EQ3 are equivalent

EQ2:

$$\min_{w,\xi} \frac{\lambda}{2} \|w\|^2 + \sum_{i=1}^{T} \xi_i$$

$$\xi_i \geq 0$$

$$y_i w^T f_i \geq 1 - \xi_i$$

EQ3:

$$\min_{w} \frac{\lambda}{2} \|w\|^2 + \sum_{i=1}^{T} \max(0, 1 - y_i w^T f_i)$$

Proof:

$$\xi_i \geq 1 - y_i w^T f_i$$

and

$$\xi_i \geq 0$$

So:

$$\xi_i = \max(0, 1 - y_i w^T f_i)$$

Plugging in:

$$\boxed{\min_{w,\xi} \frac{\lambda}{2} \|w\|^2 + \sum_{i=1}^{T} \xi_i = \min_{w} \frac{\lambda}{2} \|w\|^2 + \sum_{i=1}^{T} \max(0, 1 - y_i w^T f_i)}$$

2.2.2) Prove EQ4 is a Convex Optimization Problem
EQ4:

$$\min_{w} \sum_{i=1}^{T} \frac{\lambda}{2T} \|w\|^2 + \max(0, 1 - y_i w^T f_i)$$

Where $\frac{\lambda}{2T} \|w\|^2$ is convex, $\max(0, 1 - y_i w^T f_i)$ is affine, and the sum of a convex function and an affine function is convex.

So $\frac{\lambda}{2T} \|w\|^2 + \max(0, 1 - y_i w^T f_i)$ is convex for every i. The sum of a convex function with another convex function is also convex.

This means $\sum_{i=1}^{T} \frac{\lambda}{2T} \|w\|^2 + \max(0, 1 - y_i w^T f_i)$ is a convex function.

So $\min_w \sum_{i=1}^{T} \frac{\lambda}{2T} \|w\|^2 + \max(0, 1 - y_i w^T f_i)$ is a convex optimization problem.

2.2.3) Prove the valid sub-gradient for EQ5
EQ5

$$l_t(w) = \frac{\lambda}{2T} \|w\|^2 + \max(0, 1 - y_t w^T f_t)$$

if $1 - y_t w^T f_t > 0$

$$l_t(w) = \frac{\lambda}{2T} \|w\|^2 + 1 - y_t w^T f_t$$

$$\nabla l_t(w) = \frac{\lambda}{T} w - y_t f_t$$

else

$$l_t(w) = \frac{\lambda}{2T} \|w\|^2$$

$$\nabla l_t(w) = \frac{\lambda}{T} w$$

So together:

$$\boxed{\nabla l_t(w) = \begin{cases} \frac{\lambda}{T} w - y_t f_t & 1 - y_t w^T f_t > 0 \\ \frac{\lambda}{T} w & otherwise \end{cases}}$$

2.3) Implementation
Datasets Selected:
A) Ground: ID = 1200, Number of Points = 67161
B) Facade: ID = 1400, Number of Points = 12092
Total Misclassification Error: 417 Points mislabeled, or 0.53% of the total data.
Ground Point Misclassification Error: 4 Points mislabeled or 0.01% of the ground data.
Facade Point Misclassification Error: 413 Points mislabeled or 3.42% of the facade data.

Both sets were classified pretty well, this is probably due to the fact that both the ground and the facade are flat and easy to identify, with the ground being easier than the facade.

SVM went through and learned to classify ground and facade, by going through 67161 ground points, 12092 facade points. Each point has 9 different features associated with it. This entire process took 4.28 seconds.