Q1

$$w_{good} = (0,0,0)$$
$$w_{bad} = (0,0,0)$$
$$w_{ugly} = (0,0,0)$$

Processing example #1:

$$\text{SCORE}(good) = w_{good} \cdot x_1$$
$$= (0,0,0)\cdot(0,1,0)$$
$$= 0$$

$$\text{SCORE}(bad) = w_{bad} \cdot x_1$$
$$= 0$$

$$\text{SCORE}(ugly) = w_{ugly} \cdot x_1$$
$$= 0$$

Prediction = Highest scoring label (Break ties randomly) → Lexicographic order: bad > good > ugly

= bad.

$\boxed{\text{Error}}$

$$w_{good} = w_{good} + x_1$$
$$= (0,0,0) + (0,1,0)$$
$$= (0,1,0)$$

$$w_{bad} = w_{bad} - x_1$$
$$= (0,-1,0)$$

Processing example #2:

$$\text{SCORE}(\text{good}) = w_{good} \cdot x_2$$
$$= (0,1,0) \cdot (1,0,1)$$
$$= 0$$

$$\text{SCORE}(\text{bad}) = w_{bad} \cdot x_2$$
$$= (0,-1,0) \cdot (1,0,1)$$
$$= 0$$

$$\text{SCORE}(\text{ugly}) = w_{ugly} \cdot x_2$$
$$= 0$$

Prediction = highest scoring label (break ties)
= bad

$\boxed{\text{NO ERROR}}$

$\Rightarrow$ no change in weights.

Processing example #3:

$$\text{SCORE}(\text{good}) = w_{good} \cdot x_3$$
$$= (0,1,0) \cdot (1,1,1)$$
$$= 1$$

$$\text{SCORE}(\text{bad}) = w_{bad} \cdot x_3$$
$$= (0,-1,0) \cdot (1,1,1)$$
$$= -1$$

$$\text{SCORE}(\text{ugly}) = w_{ugly} \cdot x_3$$
$$= (0,0,0) \cdot (1,1,1)$$
$$= 0$$

Prediction = Highest scoring label
= good.

$\boxed{\text{ERROR}}$

$$w_{ugly} = w_{ugly} + x_3$$
$$= (0,0,0) + (1,1,1)$$
$$= (1,1,1)$$

$$w_{good} = w_{good} - x_3$$
$$= (0,1,0) - (1,1,1)$$
$$= (-1,0,-1)$$

Processing Example #4:

$$\text{SCORE}(\text{good}) = w_{good} \cdot x_4$$
$$= (-1,0,-1) \cdot (1,0,0)$$
$$= -1$$

$$\text{SCORE}(\text{bad}) = w_{bad} \cdot x_4$$
$$= (0,-1,0) \cdot (1,0,0)$$
$$= 0$$

$$\text{SCORE}(\text{ugly}) = w_{ugly} \cdot x_4$$
$$= (1,1,1) \cdot (1,0,0)$$
$$= 1$$

Prediction = Highest scoring label
= ugly.

$\boxed{\text{ERROR}}$

$$w_{bad} = w_{bad} + x_4$$
$$= (0,-1,0) + (1,0,0)$$
$$= (1,-1,0)$$

$$w_{ugly} = w_{ugly} - x_4$$
$$= (1,1,1) - (1,0,0)$$
$$= (0,1,1)$$

Processing Example #5:

Same procedure as above.

Q2

$$F(x) = \text{Sign}\left(\sum_{i=1}^{n} \alpha_i K(x_i, x) \right)$$

n = no of training examples.

$\alpha_1, \alpha_2, \ldots, \alpha_n$ = dual weights

$K(x_i, x)$ = Kernel function between inputs x_i and x

x = input example for which we want to make prediction.

Linear Kernel : $K(x, x') = x \cdot x'$
(dot product)

Polynomial kernel with degree 3 :

$$K(x, x') = (1 + x \cdot x')^3$$

Say,

good = +1
bad = −1
} Mapping of binary labels.

(a) Linear Kenel

$\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$, $\alpha_4 = 0$, $\alpha_5 = 0$

Processing input example #1:

$$F(x_1) = \text{Sign}\left(\sum_{i=1}^{5} 0 * K(x_i, x_1)\right)$$

$$= \text{Sign}(0)$$

[Say $> 0 \Rightarrow +ve$
$\leq 0 \Rightarrow -ve$

Prediction = $-ve$ (bad)

$\boxed{\text{ERROR}}$

$$\alpha_1 = \alpha_1 + y_1$$
$$= \alpha_1 + 1$$
$$= 0 + 1 = 1$$

Processing input example #2:

$$F(x_2) = \text{Sign}\left(\alpha_1 * K(x_1, x_2) + 0 + 0 + 0 + 0\right)$$

$$= \text{Sign}\left(1 * K(x_1, x_2)\right)$$

$\Downarrow$

$$(0,1,0) \cdot (1,0,1)$$
$$= 0$$

$= \text{sign}(1 * 0)$

$= \text{sign}(0)$

Prediction = −ve (bad)

$\boxed{\text{NO ERROR}}$

Processing input example # 3 :

$$F(x_3) = \text{sign}(\alpha_1 * K(x_1, x_3) + 0 + 0 + 0 + 0)$$

$$= \text{sign}(1 * \underbrace{K(x_1, x_3)}_{\substack{\Downarrow \\ (0,1,0)\cdot(1,1,1) \\ = 1}})$$

$= \text{sign}(1 * 1)$

Prediction = +ve (good)

$\boxed{\text{NO ERROR}}$

Processing input example #4 :

$$F(x_4) = \text{Sign}\,(\alpha_1 * K(x_1, x_4) + 0 + 0 + 0 + 0)$$

$$= \text{Sign}\,(1 * \underbrace{K(x_1, x_4)}_{\substack{(0,1,0)\ (1,0,0) \\ = 0}})$$

$$= \text{Sign}\,(0)$$

Prediction = −ve (bad)

$\boxed{\text{NO ERROR}}$

Processing input example #5 :

$$F(x_5) = \text{Sign}\,(\alpha_1 * K(x_1, x_5) + 0 + 0 + 0 + 0)$$

$$= \text{Sign}\,(1 * \underbrace{K(x_1, x_5)}_{\substack{(0,1,0)\cdot(0,0,1) \\ = 0}})$$

$$= \text{Sign}\,(0)$$

Prediction = −ve (bad)

$\boxed{\text{ERROR}}$

$$\alpha_5 = \alpha_5 + y_5$$
$$= 0 + 1$$
$$= 1$$

Dual weights:

$\alpha_1 = 1$, $\alpha_2 = 0$, $\alpha_3 = 0$, $\alpha_4 = 0$, $\alpha_5 = 1$

(b) polynomial kernel width degree = 3

Same as above but use the kernel definition

$$K(x, x') = (1 + x \cdot x')^3$$

Q3

To reduce the computational complexity of nearest neighbor classifier using the given property, we will do the following:

- Pre-compute the scaled mean of each input example from training data

- Given a test point (example), we first compute its scaled mean and then compute its distance to the pre-computed training data

- Since distance function is computed in one-dimension (1D) instead of the original dimensionality ($d > 1$), the overall computational complexity is reduced.

Q4 Yes, of course.

Here is a simple construction for binary features and binary class labels, but it can be generalized easily.

Any classifier over binary features can be thought of as a boolean function with the binary label as the output. ~~We will show~~

We will show that a decision tree can encode any truth-table, and therfore, any boolean function. So it can encode any set of "consistent" rules to produce an equivalent classifier.

Consider the following decision tree construction. The node at the root is "feature 1", the nodes at the next level correspond to "feature 2" and so on until the last level of the non-leaf nodes, which correspond to "feature n".

There will be 2^n leaves corresponding to every possible input values for the "n" features (one possible row in the truth table)

For each leaf, just run the rule classifier on the input corresponding to the path to the leaf and store the output of the rule classifier as the label of the leaf.

This mimics the rule classifier (set of rules) on all possible inputs by construction.
Therefore, they both are equivalent!

Midterm Exam Solution for Q15:

(a) SCORE (good)

$= w_{good} \cdot x$

$= (-1, -1, -1, -1) \cdot (-1, +1, +1, +1)$

$= (-1 * -1) + (-1 * +1) + (-1 * +1) + (-1 * +1)$

$= 1 - 1 - 1 - 1$

$= -2$

SCORE (bad)

$= w_{bad} \cdot x$

$= (-1, +1, +1, -1) \cdot (-1, +1, +1, +1)$

$= 1 + 1 + 1 - 1$

$= 2$

SCORE (ugly)

$= w_{ugly} \cdot x$

$= (-1, -1, -1, -1) \cdot (-1, +1, +1, +1)$

$= 1 - 1 - 1 - 1 = -2$

Prediction $\hat{y}$ = Highest scoring label

= bad.

Not same as correct label "good"

$\boxed{\text{ERROR}}$

$$w_{good} = w_{good} + x$$

$$= \cancel{(-2, 0, 0, 0)}$$

$$(-1, -1, -1, -1) + (-1, +1, +1, +1)$$

$$= (-2, 0, 0, 0)$$

$$w_{bad} = w_{bad} - x$$

$$= (-1, +1, +1, -1)$$

$$- (-1, +1, +1, +1)$$

$$= (0, 0, 0, -2)$$

$w_{ugly} \rightarrow$ unchanged