CS 6362 Machine Learning, Fall 2017: Homework 2

Venkataramana Nagarajan

Question 1:

Question 2:

Question 3:

(a)

(b)

Question 4:

(a)

(b)

(c)

Question 5:

(a)

(b)

Question 2: Since the cross-validation error is the same for both set of parameters. This question can be answered with the following reasoning:

- It is dependant how many more support vectors are being generated by the second set of parameters. Higher number of support vectors lead for a higher chance of over-fitting.

- The computational time involved in prediction for these case of parameters, it is known that predictions will be computationally faster with lesser number of support vectors. Hence they would be preferable.

- An analogy can also be taken from Occam's razor that the simpler solution is always preferable as these vectors can be considered as generated assumptions by the model for giving predictions. This would again lead to a choice of parameters which provided lesser number of support vectors.

Question 3:

Question 4: a) In a polynomial kernel increasing 'd' which is the number of dimensions will increase the chances of over-fitting. The reason being that as 'd' is increased the number of features present in the higher dimensional space will increase drastically. Higher dimension projection of the existing feature space not only make the model overfit but also they make it computationally expensive to perform. Hence this kernel suffers from the curse of dimensionality.

b) The Gaussian kernel is not directly affected by increasing dimensions. While the sigma governs the estimation for bias and variance. The affect of sigma is not directly relatable to curse of dimensionality because the expansion of this kernel is a taylor series and that is infinite all scenarios eitherways. Hence increasing or decreasing sigma will not be a cause for over-fitting however it can be used to control scenarios of bias and variance for the model.

c) Using the assumption that the feature maps are not infinite

Since:

$$K(x_i, x_i') = <\phi(x_i), \phi(x_i')> \tag{1}$$

Since H(a) = max(1-a,0) it can be inferred that H(a) ¿= 0 for all scenarios.
If 1-a¿0 then:

$$a = y_i(w^t x_i) \tag{2}$$

This is representative of a linear function. If we suppose a scenario for two functions are present in a max function and it can be proven that the max is a convex then it can be said the same for the function above.

Consider two generic functions as h = max(f,g)
Linear inequality for convexity:

$$h(z) <= th(x) + (1t)h(y)h(z) <= th(x) + (1t)h(y)where, z = tx + (1t)yz = tx + (1t)y \tag{3}$$

Since,

$$h = max f, g \tag{4}$$

This is equivalent to two inequalities:

$$f(z) <= th(x) + (1t)h(y) \tag{5}$$

$$g(z) <= th(x) + (1t)h(y) \tag{6}$$

By convexity of f one knows that:

$$f(z) <= tf(x) + (1t)f(y)f(z) <= tf(x) + (1t)f(y) \tag{7}$$

Therefore we can say that H(a) is a convex function of a.