HW7 Solution, ECE7252 Spring 2008 (March 26)

1. Exercise 6.2.

For a local polynomial regression, the Taylor series expansion

$$\hat{f}(x_0) = \sum_{j=0}^{\infty} f^{(j)}(x_0) \frac{1}{j!} \sum_{i=1}^{N} (x_i - x_0)^j \, l_i(x_0)$$

still has to satisfy perfect approximation at the expansion point x_0, thus for a degree k polynomial regression:

$$\sum_{i=1}^{N} l_i(x_0) = \sum_{i=1}^{N} (x_i - x_0)^0 \, l_i(x_0) 1$$

$$\sum_{i=1}^{N} (x_i - x_0)^j \, l_i(x_0) = 0, \; j = 1, 2, \cdots, k$$

$$f^{(n)}(x_0) = 0, \; n \geq k+1$$

2 Analysis of Kernel Methods

Suppose we have a linear function $y = f(x)$ applied to data that is uniformly spread over the interval $[0, 10]$, but we fit a kernel regression function such as the Nadaraya-Watson kernel-weighted average in Equation (6.2). Which part of the interval $[0, 10]$ will be most prone to error? Why?

The end portions of the interval will be most prone to error. The kernel regression function takes into account local contributions of other data points to perform regression on a given input point. For those points toward the center of the interval, its neighbors to both the left and right will contribute to the regression function, giving an accurate description of the behavior of the data. However, for the points closer to either end of the interval (closer to 0 or 10), there will be a dearth of neighbors on one side; for numbers close to 0, there many fewer neighbors to the left (more negative) that will contribute to the regression function, while for the points closer to 10, there will be less neighbors to the right (more positive) to contribute to the regression function. As a result, the behavior of the true function is not entirely captured at these points.

Since the function we are dealing with is linear, and the data is uniformly spread over the interval, the resulting y points will be linearly spread on the interval $[f(0), f(10)]$. Since this function is linear, for all points close enough to $f(5)$ such that there are an equal number of neighboring points on either side to contribute to the regression, the averaging operations will result in a proper output. Now consider conditions at either end, and assume without loss of generality that $f(0) < f(10)$ (the function is linear and hence monotonically increasing if this is the case). Then for a given point x_L near 0, after applying the kernel, there will be a greater contribution from points $x_o > x_L$ than from points $x_i < x_L$ because there is a limitation on the number of points x_i. As a result, the estimates will be greater than the true values, since $x_o > x_i \Rightarrow f(x_0) > f(x_i)$. Similarly, the estimates for points x_H near 10, after applying the kernel, will be lesser than the true values, since there will be fewer $x_0 > x_H$ and more $x_i < x_H$.

Therefore, we can conclude that the error will increase as the distance $|x - 5|$ increases from 0 to 5 (*i.e.*as x moves from 5 to either 10 or 0).

3. A Gaussian Mixture Model has the following distribution.

$$p(X) = \sum_{k=1}^{K} \omega_k \cdot N(X \mid \mu_k, \Sigma_k)$$

Here, we can directly compute the expectation.

$$\mu = E[X] = \int X \sum_{k=1}^{K} \omega_k \cdot N(X \mid \mu_k, \Sigma_k) dx$$

$$= \sum_{k=1}^{K} \omega_k \cdot \int X \cdot N(X \mid \mu_k, \Sigma_k) dx \; = \sum_{k=1}^{K} \omega_k \mu_k$$

$$\Sigma = E[XX^T] - \mu\mu^T$$

$$= \sum_{k=1}^{K} \omega_k (\Sigma_k + \mu_k \mu_k^T) - \mu\mu^T$$

$$= \sum_{k=1}^{K} \omega_k (\Sigma_k + \mu_k \mu_k^T) - \left(\sum_{k=1}^{K} \omega_k \mu_k \right) \left(\sum_{k=1}^{K} \omega_k \mu_k \right)^T$$

4. For the 3-component GMM, the weights are [1/3, 1/3, 1/3].

The centers are: [0.0, 0.3, 1.0];

The standard deviations are [0.1, 0.1, 0.2] (variances are [0.01, 0.01, 0.04])