STAT 824 hw 02

Multivariate Taylor expansion, closeness of points in high-dimensional space, Nadaraya-Watson and local polynomial estimators, CV for bandwidth selection

1. For a function $f : \mathbb{R}^d \to \mathbb{R}$, let

$$\nabla f(x_0) = \left[\begin{array}{c} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_d} f(x) \end{array} \right] \Bigg|_{x=x_0} \quad \text{and} \quad \nabla^2 f(x_0) = \left[\begin{array}{ccc} \frac{\partial^2}{\partial x_1^2} f(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_d} f(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_d \partial x_1} f(x) & \cdots & \frac{\partial^2}{\partial x_d^2} f(x) \end{array} \right] \Bigg|_{x=x_0}.$$

   For $x, x_0 \in \mathbb{R}^d$, show that

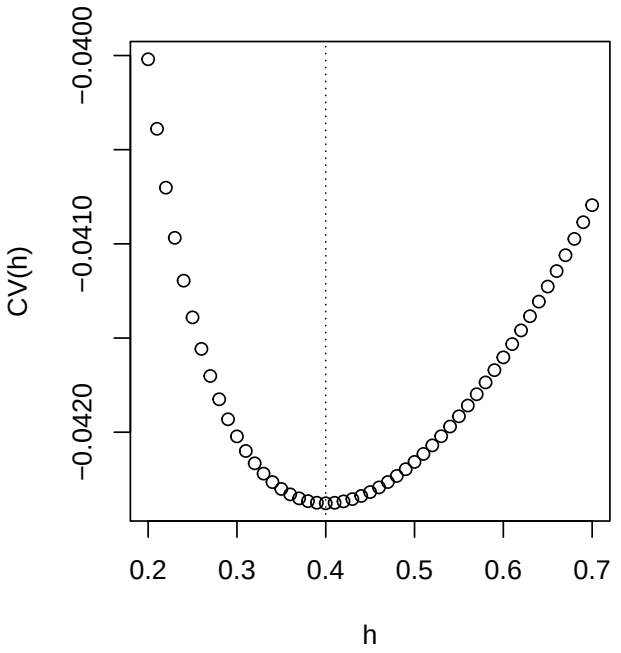
$$\sum_{|\alpha| \leq 2} \frac{D^\alpha f(x_0)}{\alpha!} (x - x_0)^\alpha = f(x_0) + [\nabla f(x_0)]^T (x - x_0) + \frac{1}{2} (x - x_0)^T [\nabla^2 f(x_0)] (x - x_0),$$

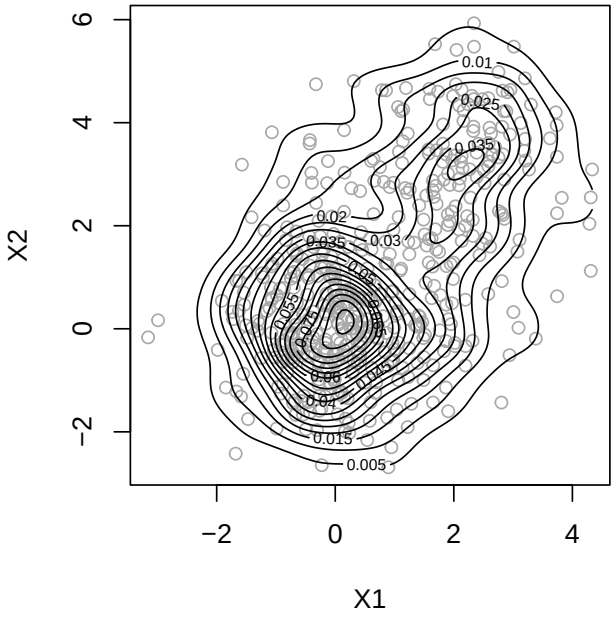
   which is the second-order Taylor expansion of f around x_0 evaluated at x.

2. Obtain $n = 500$ realizations of (X_1, X_2) by running the code

```
n <- 500; alpha <- 1/3; Z <- runif(n) < alpha; X <- matrix(NA,n,2)
X[,1] <- rnorm(n,2*Z,1);X[,2] <- rnorm(n,3*Z,1)
```

   Use leave-one-out crossvalidation to select the bandwidth for a bivariate kernel density estimator (write your own code for this). Then make a plot showing the CV criterion as a function of h and a scatterplot of your (X_1, X_2) values with contours of your estimate (at the CV choic of bandwidth) overlaid. Report your chosen bandwidth. My selected bandwidth was $\hat{h} = 0.4$. These are my plots:





3. Let $X, X_1, \ldots, X_n \in [0,1]^d$ be independent random vectors with the elements of each being independent and uniformly distributed on the interval $[0,1]$. For a vector $x \in \mathbb{R}^d$, let $\|x\|_\infty = \max_{1 \le k \le d} |x_k|$.

   (a) Show that

$$\mathbb{E} \min_{1 \le i \le n} \|X - X_i\|_\infty \ge \frac{d}{2(d+1)} \cdot \frac{1}{n^{1/d}}.$$

   (b) Give an interpretation of the claim.

4. For a set of points $(X_1, Y_1), \ldots, (X_n, Y_n)$, the Nadaraya-Watson estimator of $m(x) = \mathbb{E}[Y|X = x]$ is

$$\hat{m}_n^{\mathrm{NW}}(x) = \sum_{i=1}^n W_{ni}(x) Y_i, \quad \text{with} \quad W_{ni}(x) = \frac{K(h^{-1}(X_i - x))}{\sum_{j=1}^n K(h^{-1}(X_j - x))}.$$

   (a) Show that if $K \ge 0$ we have $\hat{m}_n^{\mathrm{NW}}(x) = \operatorname{argmin}_{\theta \in \mathbb{R}} \sum_{i=1}^n (Y_i - \theta)^2 K(h^{-1}(X_i - x))$.

   (b) Suppose $\int K(u)du = 1$ and $\int uK(u)du = 0$ and consider the kernel density estimators

$$\hat{f}_n(x,y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{Y_i - y}{h}\right) K\left(\frac{X_i - x}{h}\right) \quad \text{and} \quad \hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)$$

   of $f(x,y)$ and $f(x)$ and let $\hat{f}_n(y|x) = \hat{f}_n(x,y)/\hat{f}_n(x)$. Show that $\hat{m}_n^{\mathrm{NW}}(x) = \int y \hat{f}_n(y|x) dy$, so it is $\mathbb{E}[Y|X = x]$ taken with respect to the estimated conditional density $\hat{f}_n(y|x)$.

   (c) Show that

$$\frac{Y_i - \hat{m}_n^{\mathrm{NW}}(X_i)}{1 - W_{ni}(X_i)} = Y_i - \hat{m}_{n,-i}^{\mathrm{NW}}(X_i).$$

   (d) Explain why the fact in part (c) is useful.

5. For $n = 200$, generate data according to $Y_i = m(X_i) + \varepsilon_i$, $i = 1, \ldots, n$, where $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Beta}(1/2, 1/2)$, independent of $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{ind}}{\sim} \text{Normal}(0,1)$, where

$$m(x) = -250 \cdot (x - 1/2) \cdot \phi\left(10(x - 1/2)\right), \quad \phi(z) = (1/\sqrt{2\pi}) e^{-z^2/2}.$$

   Choose via crossvalidation a value of the bandwidth h for the local linear estimator (local polynomial of order $\ell = 1$) using ϕ as the kernel function. *Note: You will have to specify a grid of candidate h values.*

   (a) Make a plot of the function

$$\mathrm{CV}_n(h) = \frac{1}{n} \sum_{i=1}^n \left[\frac{Y_i - \hat{m}_{n,1}^{\mathrm{LP}}(X_i)}{1 - W_{ni}^*(X_i)}\right]^2$$

   over your grid of candidate bandwidths. It should dip down and rise back up. The weights $W_{ni}^*(X_i)$ are the values such that $\hat{m}_{n,1}^{\mathrm{LP}}(X_i) = \sum_{i=1}^n W_{ni}^*(X_i) Y_i$.

   (b) Make a scatterplot of the data and overlay the true function; include in the scatterplot the estimated function at your chosen value of the bandwidth.

   (c) Turn in your code.