1 The Derivative Explanation of Formulas 2.9 to 2.13

之前我们得到了在线性回归中关于 $f(x) = E[Y|X]$ 的推导。

下面我们再看 KNN 算法，KNN 算法尝试通过训练数据直接通过训练数据完成任务，对于每一个给定点 x, 我们需要对所有的输入变量 $x_i = x$ 求其对应 y_i 的均值

$$\hat{f}(x) = Ave(y_i|x_i \in Neighbor(x)) \tag{1}$$

在上式中，Ave 是求平均值的函数，$Neighbor(x)$ 是一个包含 k 个与 x 距离最近的点的领域 ($Neighbor(x)$ is a Neighbor containing the k points in T closest to x)

这里使用了两个近似

- 期望被近似为对所有样本点求均值
- 一个给定点的条件期望被”松弛”为离该点距离近的区域上的条件期望

当样本数据量很大的时候，x 周围的点会非常靠近 x. 同时，根据 mild regularity conditions, 可以证明当 $N, k- > \infty$ 同时 $\frac{k}{N} - > 0$, 我们可以得到 $\hat{f}(x) - > E[Y|X = x]$

同样对于线性回归模型

$$L(y, X\beta) = (y - X\beta)^2 \tag{2}$$

$$EPE(\beta) = \int_{x,y} (y - X\beta)^2 Pr(X = x, Y = y) dxdy \tag{3}$$

$$\begin{aligned}\frac{\partial EPE(\beta)}{\partial \beta} &= \int_{x} y - X^T(y - X\beta) Pr(X = x, Y = y) dxdy \\ &= \beta E[X^T X] - E[X^T, y]\end{aligned} \tag{4}$$

we set equation (4) to 0, we get $\beta = [E[X^T X]]^{-1} E[X^T y]$ 我们将这个表达式与之前的推导式 $\hat{\beta} = (X^T X)^{-1} X^T y$ 进行对比, 最小二乘法的解相当于用训练数据的平均值替换掉了 (4) 中的期望

所以，k-nearest neighbor 和最小二乘法都是通过求平均来近似代替条件期望，但是他们在模型的假设上具有很大的不同

- 最小二乘法假定 $f(x)$ 能被一个全局的线性函数较好的拟合
- k-nearest neighbor 假定 $f(x)$ 能被一个 locally constant function 较好的近似

如果我们使用 L_1 损失函数代替 L_2, 在这种情况下，解答是条件中位数 (conditional median)

$$\hat{f}(x) = median(Y|X = x) \tag{5}$$