EE5907/EE5027 Week 5: Non-parametric Solutions

Q1: Parzen's Window

- $p_h(x) = \frac{1}{4}\sum_{n=1}^{4} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_n - x)^2}{2}}$
- Plugging x = 2 and the data, we get

$$\begin{aligned} p_1(x=2) &= \frac{1}{4\sqrt{2\pi}} \left[e^{-\frac{(2-1)^2}{2}} + e^{-\frac{(2-3)^2}{2}} + e^{-\frac{(2-4)^2}{2}} + e^{-\frac{(2-10)^2}{2}} \right] \\ &= \frac{1}{4\sqrt{2\pi}} \left[e^{-0.5} + e^{-0.5} + e^{-2} + e^{-32} \right] \\ &= 0.134483 \end{aligned}$$

- Plugging x = 5 and the data we get

$$\begin{aligned} p_1(x=5) &= \frac{1}{4\sqrt{2\pi}} \left[e^{-\frac{(5-1)^2}{2}} + e^{-\frac{(5-3)^2}{2}} + e^{-\frac{(5-4)^2}{2}} + e^{-\frac{(5-10)^2}{2}} \right] \\ &= \frac{1}{4\sqrt{2\pi}} \left[e^{-8} + e^{-2} + e^{-0.5} + e^{-12.5} \right] \\ &= 0.0740 \end{aligned}$$

Q2: KNN

- The 3 closest datapoints for x_5 are x_1, x_2, and x_3.
- Therefore $p(y = 1|x) = 1/3$ and $p(y = 0|x) = 2/3$
- Therefore the datapoint should be classified as class 0
- The 3 closest datapoints for x_6 are x_2 (or x_1), x_3, and x_4.
- Therefore $p(y = 1|x) = 2/3$ and $p(y = 0|x) = 1/3$ (Note that x_1 and x_2 are equidistant, so I am also ok with $p(y = 1|x) = p(y = 0|x) = 1/2$)
- Therefore the datapoint should be classified as class 1