Assignment 1. Consider regression with training datasets $T_m$ of size $m$ generated as:

$$y = f(x) + \epsilon,$$  

where $\epsilon$ is the noise having $\mathbb{E}[\epsilon] = 0$ and $\text{Var}(\epsilon) = \sigma^2$. Derive bias-variance decomposition for $k$-nearest-neighbor regression. The response of the k-NN regressor is defined as:

$$h_m(x) = \frac{1}{k} \sum_{i=1}^k y_{n(x,i)} = \frac{1}{k} \sum_{i=1}^k f(x_{n(x,i)}) + \epsilon,$$

where $n(x,i)$ gives the index of $i$-th nearest neighbor of $x$ in $T_m$. For simplicity assume that all $x_i$ are the same for all training datasets $T_m$ in consideration, hence, the randomness arises from the noise $\epsilon$, only.

Give bias:

$$\mathbb{E}_x \left[ (g_m(x) - f(x))^2 \right] = \mathbb{E}_x \left[ (\mathbb{E}_{T_m}[h_m(x)] - f(x))^2 \right]$$  

and variance:

$$\text{Var}_{x,T_m}(h_m(x)).$$

Assignment 2. The output of a regression tree is defined as:

$$h(x) = \sum_{r=1}^M c_r \mathbb{I}\{x \in R_r\}$$

where $R_r$ is an input space region defined by the $r$-th tree leaf and $c_r \in \mathbb{R}$ the corresponding region’s response. The tree is trained using set $T^m = \{(x_i, y_i) \mid i = 1, \ldots, m\}$. Show that the sum of squares loss function $\sum_{i=1}^m (y_i - h(x_i))^2$ is minimized by choosing the following region responses:

$$c_r = \frac{1}{|S_r|} \sum_{x_i \in R_r} y_i$$

where $S_r = \{(x_i, y_i) : (x_i, y_i) \in T^m \land x_i \in R_r\}$.

Assignment 3. What is an optimal value of $c_r$ when the sum of absolute deviations $\sum_{i=1}^m |y_i - h(x_i)|$ is used instead of the squared loss?
Assignment 4. Bootstrapping is a method which produces $K$ datasets $T_i^m$ for $i = 1, \ldots, K$ by uniformly sampling the original dataset $T^m$ with replacement. Bootstrap datasets have typically the same size as the original dataset $|T_i^m| = |T^m| = m$. Show that as $m \to \infty$ the fraction of unique samples in $T_i^m$ approaches $1 - \frac{1}{e} \approx 63.2\%$.

Hint: apply exponential of a logarithm to a limit which emerges in a last step in order to solve it.

Assignment 5. Consider the Huber loss:

$$
\ell(y, h(x)) = \begin{cases} 
(y - h(x))^2 & \text{for } |y - h(x)| \leq \delta \\
2\delta|y - h(x)| - \delta^2 & \text{otherwise.} 
\end{cases}
$$

Define Gradient Boosting Machine using the Huber loss and discuss differences to the squared loss GBM.