Université de Lorraine, Master de mathématique, 2e année, Année 2022-2023, Semestre 1.

MSDC part 2, to understand the lecture

1. Calculate the number of models to be tested in regression when the number of variables is$p = 2$, $p = 5$, $p = 20$.
2. Slide 14, explain why $Loss(y, \hat{p}(Y = y|X)) = -2\ln(\hat{p}(Y = y|X))$ is well "a loss function
3. Slide 14 proove

$$\begin{aligned}
\mathbb{E}err(x_0) &= \mathbb{E}[(Y - \hat{h}^{(\mathcal{W})}(x_0))^2 | X = x_0] \\
&= \sigma_\varepsilon^2 + \Big(\mathbb{E}[(\hat{h}^{(\mathcal{W})}(x_0)] - h(x_0)\Big)^2 + \mathbb{E}[\Big(\hat{h}^{(\mathcal{W})}(x_0) - \mathbb{E}[\hat{h}^{(\mathcal{W})}(x_0)]\Big)^2] \\
&= \sigma_\varepsilon^2 + \text{Bias}^2(\hat{h}^{(\mathcal{W})}(x_0)) + Var(\hat{h}^{(\mathcal{W})}(x_0)) \\
&= \text{Irreducible error} + \text{Bias}^2 + \text{Variance}
\end{aligned}$$

4. Slide 14, prove formula for $\mathbb{E}err(x_0)$ in the k-nearest neighbors model
5. Slide 14, prove the result of $\mathbb{E}err(x_0)$ for the linear regression in the case p=2
6. Slide 24, give the formula of the AIC for a gaussian linear model with p covariates
7. Slide 26, give the formula of the AIC for a gaussian linear model with p covariates
8. Explain why the two last criteria lead to the same variable selection for a gaussian linear model with p covariates