

Exercise 1: Gradient Boosting (Exam 2020)

In the following, you assume that your outcome follows a \log_2 -normal distribution with density function

$$p(y|f) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log_2(y) - f)^2}{2\sigma^2}\right).$$

In other words, $\log_2(Y)$ follows a normal distribution. You observe $n = 3$ data points \mathbf{y} and want to model f using features $\mathbf{X} \in \mathbb{R}^{n \times p}$. You choose to use a gradient boosting tree algorithm.

- (a) Derive the pseudo-residuals based on the negative log-likelihood for the given distribution assumption up to all constants of proportionalities.
- (b) Given only the 3 samples $\mathbf{y} = (1, 2, 4)^\top$ and two features

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$$

explicitly derive or state with explanation

- (i) the loss-optimal initial boosting model $\hat{f}^{[0]}(\mathbf{x})$,
- (ii) the pseudo residuals $\mathbf{r}^{[1]}$ (use L_2 -loss if you haven't been able to solve (a)),
- (iii) the regression stump $R_t^{[1]}$, $t = 1, 2$,
- (iv) the boosting model $\hat{f}^{[1]}(\mathbf{x})$ as well as
- (v) the pseudo residuals $\mathbf{r}^{[2]}$

for tree base learner with depth $d = 1$ (stumps) and a step length $\nu = 1$. You are allowed to use results from the lecture. If you have not managed to derive the pseudo-residuals for the \log_2 -normal distribution, use an L_2 -loss.

- (c) What would happen in the second iteration of the previous boosting algorithm?
- (d) If you are given more data points, but still the two binary feature vectors \mathbf{x}_1 and \mathbf{x}_2 , what will happen as
 - (i) M grows
 - (ii) n grows

in terms of the model capacity (if d is kept fixed)?