1 Hinge-loss gradients (5 Points)

The function \( [z]_+ = \max(0, z) \) is called hinge. In ML, a hinge penalizes errors (when \( z > 0 \)) linearly, but raises no costs at all if \( z < 0 \).

Assume we have a single data point \( (x, y^*) \) with class label \( y^* \in \{1, \ldots, M\} \), and the discriminative function \( f(y, x) \). We penalize the discriminative function with the one-vs-all hinge loss

\[
L_{\text{hinge}}(f) = \sum_{y \neq y^*} [1 - (f(y^*, x) - f(y, x))]_+.
\]

a) What is the gradient \( \frac{\partial L_{\text{hinge}}(f)}{\partial f(y, x)} \) of the loss w.r.t. the discriminative values. For simplicity, distinguish the cases of taking the derivative w.r.t. \( f(y^*, x) \) and w.r.t. \( f(y, x) \) for \( y \neq y^* \). (3 P)

b) Now assume the parameteric model \( f(y, x) = \phi(x)^\top \beta_y \), where for every \( y \) we have different parameters \( \beta_y \in \mathbb{R}^d \). And we have a full data set \( D = \{(x_i, y_i)\}_{i=1}^n \) with class labels \( y_i \in \{1, \ldots, M\} \) and loss

\[
L_{\text{hinge}}(f) = \sum_{i=1}^n \sum_{y \neq y_i} [1 - (f(y_i, x_i) - f(y, x_i))]_+.
\]

What is the gradient \( \frac{\partial L_{\text{hinge}}(f)}{\partial \beta_y} \)? (2 P)

2 Log-likelihood gradient and Hessian (5 Points)

Consider a binary classification problem with data \( D = \{(x_i, y_i)\}_{i=1}^n \), \( x_i \in \mathbb{R}^d \) and \( y_i \in \{0, 1\} \). We define

\begin{align*}
  f(x) &= \phi(x)^\top \beta \\
  p(x) &= \sigma(f(x)), \quad \sigma(z) = 1/(1 + e^{-z}) \\
  L_{\text{all}}(\beta) &= -\sum_{i=1}^n \left[ y_i \log p(x_i) + (1 - y_i) \log[1 - p(x_i)] \right]
\end{align*}

where \( \beta \in \mathbb{R}^d \) is a vector. (NOTE: the \( p(x) \) we defined here is a short-hand for \( p(y = 1|x) \) on slide 03:9.)

a) Compute the derivative \( \frac{\partial L_{\text{all}}(\beta)}{\partial \beta} \). Tip: use the fact \( \frac{\partial}{\partial z} \sigma(z) = \sigma(z)(1 - \sigma(z)) \). (3 P)

b) Compute the 2nd derivative \( \frac{\partial^2}{\partial \beta^2} L(\beta) \). (2 P)