1 Multiple Choice 1 point

Suppose we are using squared loss $l(y, \hat{y}) = (y - \hat{y})^2$ in a regression problem with a linear prediction model: $\hat{y} = \mathbf{w}^T\mathbf{x}$.
What is the derivative of the absolute loss w.r.t. $\mathbf{w}$ for a single example with feature vector $\mathbf{x}$ and true label y?

Note: Because the absolute loss is not differentiable at zero, you can assume that y is not equal to y-hat in this problem. Note that the sign function for a real valued input x (where x is not zero) is defined as sign(x) = +1 for x > 0 and sign(x) = -1 for x < 0

- (●) $\mathbf{x}\,(\mathbf{w}^T\mathbf{x} - y)$
- ( ) $\mathbf{x}\,(y - \text{softmax}(\mathbf{w}^T\mathbf{x}))$
- ( ) $\mathbf{x}\,(\text{softmax}(\mathbf{w}^T\mathbf{x}) - y)$
- ( ) $\mathbf{x}\,(y - \mathbf{w}^T\mathbf{x})$

2 Multiple Choice 1 point

Suppose in a classification problem, the true label **y** and the predicted label **y**-hat are both hard labels. Which of the following correctly describes the nature of the cross-entropy loss in this special case?

- ( ) It is zero
- ( ) It is +infinity
- (●) It is 0 if **y** is equal to **y**-hat and +infinity otherwise
- ( ) It is 0 if **y** is equal to **y**-hat and 1 otherwise

3 Multiple Choice 1 point

Recall that cross-entropy loss l(**y**, **y**-hat) is defined as the sum of $-y_j \log(\text{y-hat}_j)$ over j. What is the partial derivative of cross-entropy loss w.r.t. y-hat_j?

- [x] $-y_j / \text{y-hat}_j$
- [ ] $\text{softmax}(\mathbf{y}\text{-hat})_j - y_j$
- [ ] $-\text{y-hat}_j / y_j$
- [ ] $\text{softmax}(\mathbf{y})_j - \mathbf{y}\text{-hat}_j$

4 Multiple Choice 1 point

Suppose **y**-hat = softmax(**o**) and I send o_1, the first component of **o**, off to **+infinity** (plus infinity) while keeping other components unchanged. What happens to $\mathbf{y}\text{-hat}_1$, the first component of **y**-hat?

- [ ] It goes to -infinity
- [ ] It goes to +infinity
- [x] It goes to 1
- [ ] It goes to 0

5 Multiple Choice 1 point

Suppose **y**-hat = softmax(**o**) and I send o_1, the first component of **o**, off to **-infinity** (minus infinity) while keeping other components unchanged. What happens to **y**-hat$_1$, the first component of **y**-hat?

- It goes to -infinity
- It goes to +infinity
- **It goes to 0**
- It goes to 1

6 Multiple Choice 1 point

Suppose I have a uniform distribution on N outcomes. What is the entropy (in bits) of this distribution?

- **$\log_2(N)$**
- $\log_2(1/N)$
- $1/N \log_2(N)$
- $1/N \log_2(1/N)$

7 Multiple Choice 1 point

Which of the following is NOT true regarding the relationship between TensorFlow and Keras?

- [x] TensorFlow offers a more a high level API (application programming interface) than Keras
- [ ] First release of Keras was before the first release of Tensorflow
- [ ] Keras was originally built on top of Theano, a precursor of TensorFlow
- [ ] Keras API occupies a front-and-center place in TensorFlow 2.0

8 Multiple Choice 1 point

What is a feature of TensorFlow that is not all supported by NumPy?

- [x] Ability to numerically compute gradients of any differentiable expression
- [ ] Ability to create vectors and matrices
- [ ] Ability to manipulate vectors and matrices
- [ ] A vast library of predefined mathematical functions such a sqrt, log, exp, etc.

9 Multiple Choice 1 point

What does the "G" in "GPU" stand for?

- [x] graphics
- [ ] grammar
- [ ] gradient
- [ ] google

10 Multiple Choice 1 point

The name "Jupyter" includes a references to 3 major programming languages for data science. Which of the following is one of them?

- (●) R
- ( ) Java
- ( ) JavaScript
- ( ) Ruby