Multiple Choice 1 point  $((y, \hat{y}) = |y - \hat{y}| = |y - w^Tx|$ 

Suppose we are using squared loss  $I(y, y-hat) = (y - y-hat)^2$  in a regression problem with a linear prediction model: v-hat =  $\mathbf{w}^{\mathsf{T}}\mathbf{x}$ .

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 v?

*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

- $x (w^T x v)$
- $\mathbf{x}$  (v softmax( $\mathbf{w}^{\mathsf{T}}\mathbf{x}$ ))
- $\mathbf{x}$  (softmax( $\mathbf{w}^{\mathsf{T}}\mathbf{x}$ )  $\mathbf{v}$ )
- $\mathbf{x} (\mathbf{v} \mathbf{w}^{\mathsf{T}} \mathbf{x})$

- Vw ly-wxl
  - $= -x \cdot (sign(w^{7}x-b))$

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

- [= \( \frac{1}{2} \frac{1}{2} \right| \frac{1}{2} \frac{1}{2} \]
- 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
- $y_{i}=1, y_{i}=0 \implies -(-\infty)$

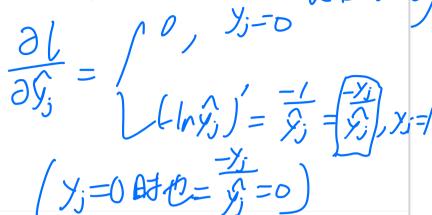
那么为一般 约二1,分地方分二0一种

因而2要1/a地(不相同,一定在1=100;否则(=0

 $|(y,\hat{y})| = \xi - \chi_i |n \hat{y}_i| = -|n \hat{y}_{\alpha}|$ 

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

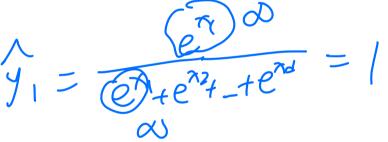
- -y<sub>i</sub> / y-hat<sub>i</sub>
- softmax(y-hat)<sub>j</sub> y<sub>j</sub>
- -y-hat<sub>j</sub>/y<sub>j</sub>
- softmax(y)<sub>j</sub> y-hat<sub>j</sub>



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 y-hat<sub>1</sub>, the first component of y-hat?

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



Suppose **y**-hat = softmax( $\mathbf{o}$ ) and I send o<sub>1</sub>, the first component of  $\mathbf{o}$ , off to -infinity minus infinity) while keeping other components unchanged. What happens to  $\mathbf{y}$ -hat<sub>1</sub>, the first component of  $\mathbf{y}$ -hat?

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

Multiple Choice 1 point

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

- $\log_2(1/N)$
- 1/N log<sub>2</sub>(N)
- 1/N log<sub>2</sub>(1/N)

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

0	R
	Java
	JavaScript
	Ruby