

$$\frac{1 + e^{x_1 m_1 + x_2 m_2 + \gamma}}{e^{x_1 m_1 + x_2 m_2}} =$$

$$\frac{\partial}{\partial t} = \frac{q \rho}{\sigma - x_1 m_1 - x_2 m_2}$$

$$c) \quad \frac{1}{1 + e^{-(x_1 m_1 + x_2 m_2 + \gamma)}}$$

$$X_1 \frac{(1 + e^{-x_1 m_1 - x_2 m_2 - \sigma})}{(1 + e^{-x_1 m_1 - x_2 m_2 - \sigma})^2}$$

$$+ \frac{e^{-x_1 m_1 - x_2 m_2 - \sigma}}{(1 + e^{-x_1 m_1 - x_2 m_2 - \sigma})^2}$$

$$(1 + e^{-x_1 m_1 - x_2 m_2 - \sigma})^2 (a - x_1 m_1 - x_2 m_2 - \sigma)^2$$

$$\frac{du}{d\sigma} =$$

$$b) \quad \frac{1}{(a + e^{-(x_1 m_1 + x_2 m_2 + \gamma)})}$$

- (C) Mario, Aleister, Eric
- 15 hours - 20 hours
- 3.S - 4
- 4-S

- 1) a) My sister and her boyfriend went on a break.
- either they broke up, or they went on a trip
b) I ~~saw~~ saw the old woman with the
glasses.
c) Sasha and Sophie played cards.

- 2) Statement 1: No, because for each independently
pulled set of data for each model run, the
error rate will not be the same across ~~the~~
evening runs.

Statement 2: Yes, because the trained model,
trained on T and tuned on V, will be
a more accurate, in terms, representation
of what the data looks like. By training
on one set from tuning on V, the
data are not over fit to the T set
but should be a generally good
predictor of V.

Statement 3: No, because if the more
depth is chosen randomly than the
~~depths~~ ~~will~~ error rates won't
be biased.

3) a) chain rule:

$$\begin{aligned}\frac{d\sigma}{dz} &= \frac{d}{dz} \left[\frac{1}{1+e^{-z}} \right] \\ &= \frac{d}{dz} \left(\frac{1}{1+e^{-z}} \right)^{-1} \\ &\equiv -\left(1+e^{-z} \right)^{-2} \left(-e^{-z} \right)\end{aligned}$$

$$\begin{aligned}&\equiv \frac{e^{-z}}{\left(1+e^{-z} \right)^2} \\ &= \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{\left(1+e^{-z} \right)^2} \\ &= \frac{1}{1+e^{-z}} \cdot \frac{\left(1+e^{-z} \right)^2 - 1}{\left(1+e^{-z} \right)^2} \\ &\quad \text{cancel terms} \\ &= \frac{1}{1+e^{-z}} \cdot \left(1 - \frac{1}{1+e^{-z}} \right) \\ &= \sigma(z) \cdot (1 - \sigma(z))\end{aligned}$$

Or can bound using L'Hopital's rule
and limit of Sigmoid function is 1
as z goes to infinity and 0 as z goes to negative infinity
so it is bounded between 0 and 1
using 0.5 as midpoint

V for next slides