Danil Redko

② f)

$$P(Y|X=1) = \frac{1}{1+\exp(-1)} = 0.73106\ldots$$

h) IDK geometric interpretation

j) IDK geometric interpretation

l) IDK geometric interpretation

④ a) IDK

b) IDK

c) IDK

⑤

i) Prove softmax(z') = softmax(z)

$$\text{softmax}(z') = \frac{\exp(z')}{\sum \exp(z'_j)} =$$

$$= \frac{\exp(z - \max(z))}{\exp(-\max(z)) \sum \exp(z_j)} = \frac{\cancel{\exp(-\max(z))} \exp(z)}{\cancel{\exp(-\max(z))} \sum_j \exp(z_j)} =$$

$$= \frac{\exp(z)}{\sum_j \exp(z_j)} = \text{softmax}(z) \text{ as needed}$$

It does not cause over flow since all numbers in $\vec{z}$ are now non-positive. This cannot lead to overflow because computer would have enough bits to compute this. However, it may lead to underflow.

ii) $\log\left(\frac{\exp(z)}{\sum_j \exp(z_j)}\right) = \log(\exp(z)) - \log\left(\sum_j \exp(z_j)\right) =$

$\ln(\exp(x)) = x$

$$= z - \log\left(\sum_j \exp(z_j)\right) = \log(y_k)$$

It does not cause underflow since we do not have to compute exp(z) which may result in extremely small number computer cannot process.

7 IDK