Task 1

a) $\hat{\vec{y}}$ has dimensions of $p \times 1$. Each row represents a probability

b) The model will output equal probabilities for all classes

Task 2

a) 1. elementwise exponentiation of input

2. sum over all elements, keepdims ensures that resulting sum has same dims as input

3. calculate class probs

b) $\text{softmax}(\vec{z}) = \frac{e^{z}}{\sum_{j=1}^{n} e^{z_j}} \;;= \text{softmax}(\vec{z} - \alpha \cdot (1, \ldots, 1)^T) = \frac{e^{(z-\alpha)}}{\sum_{j=1}^{n} e^{(z_j - \alpha)}}$

c) overflow for large numbers

underflow " small "

z.B. substract max value from all elements

Task 3

a) When the model predicts the correct true class with probability 1

b) $L^{(i)} = -\sum_{j=1}^{n} y_j^{(i)} \log(\hat{y}_j^{(i)})$ $\qquad \frac{\partial L^{(i)}}{\partial o_j^{(i)}} = \hat{y}_j^{(i)} - y_j^{(i)}$

$$\frac{\partial L^{(i)}}{\partial \hat{y}_j^{(i)}} = -\frac{y_j^{(i)}}{\hat{y}_j^{(i)}} \qquad \frac{\partial \hat{y}_j^{(i)}}{\partial o_j^{(i)}} = \hat{y}_j^{(i)}(1 - \hat{y}_j^{(i)})$$

$$\frac{\partial o_j^{(i)}}{\partial o_j^{(i)}} = 1$$

$$\frac{\partial L^{(i)}}{\partial o_j^{(i)}} = \frac{\partial L^{(i)}}{\partial \hat{y}_j^{(i)}} \cdot \frac{\partial \hat{y}_j^{(i)}}{\partial o_j^{(i)}}$$

$$= -\frac{y_j^{(i)}}{\hat{y}_j^{(i)}} \cdot \frac{\partial \hat{y}_j^{(i)}}{\partial o_j^{(i)}}$$

$$= -\frac{y_j^{(i)}}{\hat{y}_j^{(i)}} \cdot \hat{y}_j^{(i)}(1 - \hat{y}_j^{(i)})$$

$$= -y_j^{(i)}(1-\hat{y}_j^{(i)})$$
$$= \hat{y}_j^{(i)} - y_j^{(i)}$$