Softmax regression

$C = \#classes,\ n^{[L]} = C,\ \hat{y}\ is\ (C, 1)$

$$\hat{y} \rightarrow \begin{cases} P(class1|x) \\ P(class2|x) \\ \cdots \\ P(classC|x) \end{cases}$$

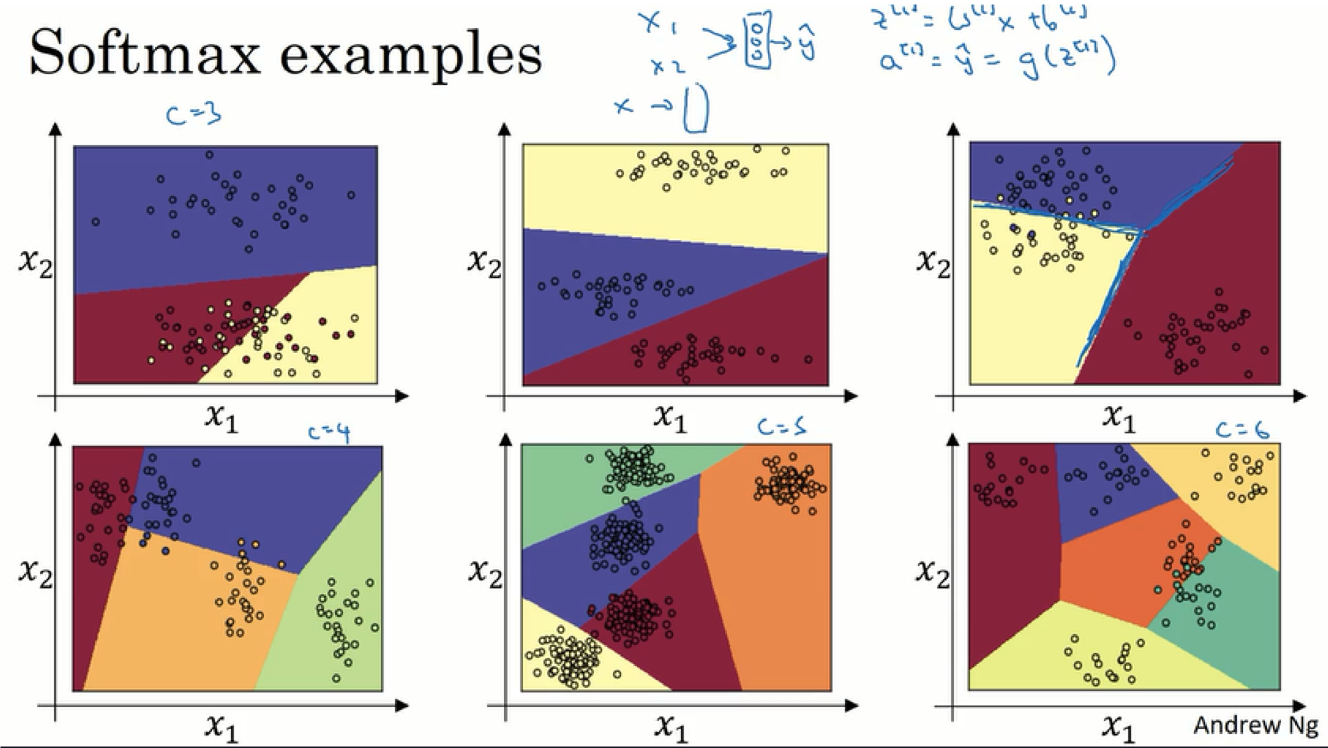
$z^{[L]} = W^{[L]} a^{[L-1]} + b^{[L]}$

$Activation\ function : g^{[L]}(z^{[L]})$

$t = e^{(z^{[L]})}$

$a^{[L]} = \frac{e^{z^{[L]}}}{\sum_i t_i},\ a_i^{[L]} = \frac{t_i}{\sum_i t_i}$

$a^{[L]} = g^{[L]}(z^{[L]}),\ a^{[L]}\ and\ z^{[L]}\ are\ (C, 1)$



---

training a softmax classifier

- **loss function**

  $\mathscr{L}(\hat{y}, y) = -\sum_{j=1}^{C} y_j log(\hat{y}_j)$

  it looks at wherever the ground true class in the training set is, and tries to make the corresponding probabilities as high as possible
- $\mathscr{J}(W^{[1]}, b^{[1]}, \cdots) = \frac{1}{m} \sum_i \mathscr{L}(\hat{y}^{(i)}, y^{(i)})$
- $Y = [\, y^{(i)} \quad \cdots \quad y^{(m)} \,],\ Y \in \mathbb{R}^{(C,m)}$

- $\hat{Y} = \begin{bmatrix} \hat{y}^{(i)} & \cdots & \hat{y}^{(m)} \end{bmatrix},\ \hat{Y} \in \mathbb{R}^{(C,m)}$
- **Backprop**
  - $\frac{\partial \mathscr{J}}{\partial z^{[L]}} = dz^{[L]} = \hat{y} - y,\ dz^{[L]} \in \mathbb{R}^{(C,1)}$