$$\mathbf{h} = \mathbf{w}^T\mathbf{X}$$

$$\text{Logistic regression: } \mathbf{z} = \sigma(\mathbf{h}) = \frac{1}{1+e^{-\mathbf{h}}}$$

$$\text{Cross-entropy loss: } J(\mathbf{w}) = -(\mathbf{y}log(\mathbf{z}) + (1-\mathbf{y})log(1-\mathbf{z}))$$

$$\text{Use chain rule: } \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{\partial J(\mathbf{w})}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{w}}$$

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{z}} = -(\frac{\mathbf{y}}{\mathbf{z}} - \frac{1-\mathbf{y}}{1-\mathbf{z}}) = \frac{\mathbf{z}-\mathbf{y}}{\mathbf{z}(1-\mathbf{z})}$$

$$\frac{\partial \mathbf{z}}{\partial \mathbf{h}} = \mathbf{z}(1-\mathbf{z})$$

$$\frac{\partial \mathbf{h}}{\partial \mathbf{w}} = \mathbf{X}$$

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \mathbf{X}^T(\mathbf{z}-\mathbf{y})$$

$$\text{Gradient descent: } \mathbf{w} = \mathbf{w} - \alpha\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}}$$