1 Linear regression

- $m \in \mathbb{N}$, number of examples.
- $n \in \mathbb{N}$, example size.
- $A \bowtie B$, element-wise using broadcasting (python like).

$$\mathbf{X} \in \mathbb{R}^{m\times n}$$
$$\mathbf{x}_i \in \mathbb{R}^{1\times n}$$
$$\mathbf{y} \in \mathbb{R}^{m\times 1}$$

$$\mathbf{w} \in \mathbb{R}^{1\times n}$$
$$b \in \mathbb{R}$$

1.1 Cost function

$$J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \frac{1}{m}\sum_{i=1}^{m}\left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right)^2$$

$$J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \frac{1}{m}\sum_{i=1}^{m}\left(\mathbf{w}\mathbf{x}_i^\top\mathbf{x}_i\mathbf{w}^\top + 2b\mathbf{w}\mathbf{x}_i^\top + b^2 - 2\mathbf{w}\mathbf{x}_i^\top y_i - 2by_i + y_i^2\right)$$

$$J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \mathbf{w}\left[\frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i^\top\mathbf{x}_i\right]\mathbf{w}^\top + 2b\mathbf{w}\frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i^\top + \frac{1}{m}\sum_{i=1}^{m}b^2 - 2\mathbf{w}\frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i^\top y_i - 2b\frac{1}{m}\sum_{i=1}^{m}y_i + \frac{1}{m}\sum_{i=1}^{m}y_i^2$$

$$\boxed{J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \mathbf{w}\left[\frac{1}{m}\mathbf{X}^\top\mathbf{X}\right]\mathbf{w}^\top + 2b\mathbf{w}\boldsymbol{\mu}_2^\top\left(\mathbf{X}\right) + b^2 - 2\mathbf{w}\boldsymbol{\mu}_2^\top\left(\mathbf{X}\bowtie\mathbf{y}\right) - 2b\mu\left(\mathbf{y}\right) + \frac{1}{m}\mathbf{y}^\top\mathbf{y}}$$

1.2 Gradient (weights)

$$\frac{\partial}{\partial w^{(j)}}J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial w^{(j)}}\left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right)^2$$

$$\frac{\partial}{\partial w^{(j)}}J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \frac{2}{m}\sum_{i=1}^{m}\left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right)\frac{\partial}{\partial w^{(j)}}\left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right)$$

$$\frac{\partial}{\partial w^{(j)}}J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \frac{2}{m}\sum_{i=1}^{m}\left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right)\left(\frac{\partial}{\partial w^{(j)}}\mathbf{w}\mathbf{x}_i^\top + \frac{\partial}{\partial w^{(j)}}b - \frac{\partial}{\partial w^{(j)}}y_i\right)$$

$$\frac{\partial}{\partial w^{(j)}}J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \frac{2}{m}\sum_{i=1}^{m}\left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right)\left(\frac{\partial}{\partial w^{(j)}}\mathbf{w}\mathbf{x}_i^\top\right)$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{2}{m} \sum_{i=1}^{m} \left(\mathbf{w}\mathbf{x}_i^\top + b - y_i\right) x_i^{(j)}$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \mathbf{w}\frac{2}{m} \sum_{i=1}^{m} \mathbf{x}_i^\top x_i^{(j)} + b\frac{2}{m} \sum_{i=1}^{m} x_i^{(j)} - \frac{2}{m} \sum_{i=1}^{m} y_i x_i^{(j)}$$

$$\frac{\partial}{\partial \mathbf{w}} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{2}{m} \begin{bmatrix} \mathbf{w}\sum_{i=1}^{m} \mathbf{x}_i^\top x_i^{(1)} \\ \mathbf{w}\sum_{i=1}^{m} \mathbf{x}_i^\top x_i^{(2)} \\ \vdots \\ \mathbf{w}\sum_{i=1}^{m} \mathbf{x}_i^\top x_i^{(n)} \end{bmatrix} + \frac{2b}{m} \begin{bmatrix} \sum_{i=1}^{m} x_i^{(1)} \\ \sum_{i=1}^{m} x_i^{(2)} \\ \vdots \\ \sum_{i=1}^{m} x_i^{(n)} \end{bmatrix} - \frac{2}{m} \begin{bmatrix} \sum_{i=1}^{m} y_i x_i^{(1)} \\ \sum_{i=1}^{m} y_i x_i^{(2)} \\ \vdots \\ \sum_{i=1}^{m} y_i x_i^{(n)} \end{bmatrix}$$

$$\frac{\partial}{\partial \mathbf{w}} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{2}{m}\mathbf{w} \sum_{i=1}^{m} \mathbf{x}_i^\top \mathbf{x}_i + \frac{2b}{m} \sum_{i=1}^{m} \mathbf{x}_i - \frac{2}{m} \sum_{i=1}^{m} y_i \mathbf{x}_i$$

$$\boxed{\frac{\partial}{\partial \mathbf{w}} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = 2\left(\mathbf{w}\left[\frac{1}{m}\mathbf{X}^\top \mathbf{X}\right] + b\boldsymbol{\mu}_2\left(\mathbf{X}\right) - \boldsymbol{\mu}_2\left(\mathbf{X} \bowtie \mathbf{y}\right)\right)}$$

1.3 Gradient (bias)

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{1}{m} \sum_{i=1}^{m} \frac{\partial}{\partial b} \left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right)^2$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{2}{m} \sum_{i=1}^{m} \left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right) \frac{\partial}{\partial b} \left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right)$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{2}{m} \sum_{i=1}^{m} \left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right) \left(\frac{\partial}{\partial b}\mathbf{w}\mathbf{x}_i^\top + \frac{\partial}{\partial b} b - \frac{\partial}{\partial b} y_i\right)$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{2}{m} \sum_{i=1}^{m} \left(\left[\mathbf{w}\mathbf{x}_i^\top + b\right] - y_i\right) \left(\frac{\partial}{\partial b} b\right)$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{2}{m} \sum_{i=1}^{m} \left(\mathbf{w}\mathbf{x}_i^\top + b - y_i\right)$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \mathbf{w}\frac{2}{m} \sum_{i=1}^{m} \mathbf{x}_i^\top + \frac{2}{m} \sum_{i=1}^{m} b - \frac{2}{m} \sum_{i=1}^{m} y_i$$

$$\boxed{\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = 2\left(\mathbf{w}\boldsymbol{\mu}_2^\top\left(\mathbf{X}\right) + b - \mu\left(\mathbf{y}\right)\right)}$$

1.4 Analitic solution

$$\frac{\partial}{\partial \mathbf{w}} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \mathbf{0} = 2\left(\mathbf{w}\left[\frac{1}{m}\mathbf{X}^\top\mathbf{X}\right] + b\boldsymbol{\mu}_2\left(\mathbf{X}\right) - \boldsymbol{\mu}_2\left(\mathbf{X} \bowtie \mathbf{y}\right)\right)$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = 0 = 2\left(\mathbf{w}\boldsymbol{\mu}_2^\top\left(\mathbf{X}\right) + b - \mu\left(\mathbf{y}\right)\right)$$

$$\begin{bmatrix} \mathbf{w}\left[\frac{1}{m}\mathbf{X}^\top\mathbf{X}\right] + b\boldsymbol{\mu}_2\left(\mathbf{X}\right) - \boldsymbol{\mu}_2\left(\mathbf{X} \bowtie \mathbf{y}\right) \\ \mathbf{w}\boldsymbol{\mu}_2^\top\left(\mathbf{X}\right) + b - \mu\left(\mathbf{y}\right) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{w}\left[\frac{1}{m}\mathbf{X}^\top\mathbf{X}\right] + b\boldsymbol{\mu}_2\left(\mathbf{X}\right) \\ \mathbf{w}\boldsymbol{\mu}_1^\top\left(\mathbf{X}\right) + b \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_2\left(\mathbf{X} \bowtie \mathbf{y}\right) \\ \mu\left(\mathbf{y}\right) \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{w} & b \end{bmatrix} \begin{bmatrix} \left[\frac{1}{m}\mathbf{X}^\top\mathbf{X}\right] & \boldsymbol{\mu}_2^\top\left(\mathbf{X}\right) \\ \boldsymbol{\mu}_2\left(\mathbf{X}\right) & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_2\left(\mathbf{X} \bowtie \mathbf{y}\right) \\ \mu\left(\mathbf{y}\right) \end{bmatrix}$$

$$\boxed{\begin{bmatrix} \mathbf{w} & b \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_2\left(\mathbf{X} \bowtie \mathbf{y}\right) \\ \mu\left(\mathbf{y}\right) \end{bmatrix} \begin{bmatrix} \left[\frac{1}{m}\mathbf{X}^\top\mathbf{X}\right] & \boldsymbol{\mu}_2^\top\left(\mathbf{X}\right) \\ \boldsymbol{\mu}_2\left(\mathbf{X}\right) & 1 \end{bmatrix}^{-1}}$$

2 Logistic regression

- $m \in \mathbb{N}$, number of examples.
- $n \in \mathbb{N}$, example size.

$$\mathbf{X} \in \mathbb{R}^{m \times n}$$

$$\mathbf{x}_i \in \mathbb{R}^{1 \times n}$$

$$\mathbf{y} \in \mathbb{B}^{m \times 1}$$

$$\mathbf{w} \in \mathbb{R}^{1 \times n}$$

$$b \in \mathbb{R}$$

2.1 Cost function

$$J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i \log\left(\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right) + \left(1 - y_i\right)\log\left(1 - \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right)$$

$$\boxed{J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m \wedge y_i = 1} \log\left(\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right) + \sum_{i=1}^{m \wedge y_i = 0} \log\left(1 - \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right]}$$

2.2 Gradient (weights)

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m\wedge y_i=1} \frac{\partial}{\partial w^{(j)}} \log\left(\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right) + \sum_{i=1}^{m\wedge y_i=0} \frac{\partial}{\partial w^{(j)}} \log\left(1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m\wedge y_i=1} \frac{1}{\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)} \frac{\partial}{\partial w^{(j)}} \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) + \sum_{i=1}^{m\wedge y_i=0} \frac{\partial}{\partial w^{(j)}} \log\left(1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m\wedge y_i=1} \frac{1}{\cancel{\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)}} \left(\cancel{\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)}\left(1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right) \frac{\partial}{\partial w^{(j)}} \left(\mathbf{w}\mathbf{x}_i^\top + b\right) + \sum_{i=1}^{m\wedge y_i=0} \frac{\partial}{\partial w^{(j)}} \log\left(1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m\wedge y_i=1} \left(x_i^{(j)} - \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)}\right) + \sum_{i=1}^{m\wedge y_i=0} \frac{\partial}{\partial w^{(j)}} \log\left(1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m\wedge y_i=1} \left(x_i^{(j)} - \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)}\right) + \sum_{i=1}^{m\wedge y_i=0} \frac{1}{1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)} \frac{\partial}{\partial w^{(j)}} \left(1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m\wedge y_i=1} \left(x_i^{(j)} - \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)}\right) - \sum_{i=1}^{m\wedge y_i=0} \frac{1}{1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)} \left(\frac{\partial}{\partial w^{(j)}} \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m\wedge y_i=1} \left(x_i^{(j)} - \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)}\right) - \sum_{i=1}^{m\wedge y_i=0} \frac{1}{\cancel{1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)}} \left(\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\cancel{\left(1-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)} \frac{\partial}{\partial w^{(j)}} \left(\mathbf{w}\mathbf{x}_i^\top + b\right)\right)\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m\wedge y_i=1} \left(x_i^{(j)} - \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)}\right) - \sum_{i=1}^{m\wedge y_i=0} \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)}\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m} y_i \left(x_i^{(j)} - \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)}\right) - \sum_{i=1}^{m} \left(1-y_i\right)\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)}\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = -\frac{1}{m}\sum_{i=1}^{m}\left[x_i^{(j)} y_i \cancel{-\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)} y_i} \cancel{+\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)} y_i} - \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)}\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \frac{1}{m}\sum_{i=1}^{m}\left[\sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) x_i^{(j)} - x_i^{(j)} y_i\right]$$

$$\frac{\partial}{\partial w^{(j)}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \frac{1}{m}\sum_{i=1}^{m} x_i^{(j)} \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) - \frac{1}{m}\sum_{i=1}^{m} x_i^{(j)} y_i$$

$$\frac{\partial}{\partial \mathbf{w}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \frac{1}{m}\sum_{i=1}^{m} x_i^{(j)} \sigma\left(\mathbf{w}\mathbf{x}_i^\top + b\right) - \boldsymbol{\mu}_1\left(\mathbf{X} \bowtie \mathbf{y}\right)$$

$$\boxed{\frac{\partial}{\partial \mathbf{w}} J\left(\mathbf{X},\mathbf{y},\mathbf{w},b\right) = \boldsymbol{\mu}_1\left(\mathbf{X} \bowtie \boldsymbol{\sigma}\left(\mathbf{X}\mathbf{w}^\top + \mathbf{b}\right)\right) - \boldsymbol{\mu}_1\left(\mathbf{X} \bowtie \mathbf{y}\right)}$$

2.3 Gradient (bias)

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m \wedge y_i = 1} \frac{\partial}{\partial b} \log\left(\sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)\right) + \sum_{i=1}^{m \wedge y_i = 0} \frac{\partial}{\partial b} \log\left(1 - \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)\right)\right]$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m \wedge y_i = 1} \frac{1}{\sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)} \frac{\partial}{\partial b} \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right) + \sum_{i=1}^{m \wedge y_i = 0} \frac{1}{1 - \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)} \frac{\partial}{\partial b}\left(1 - \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)\right)\right]$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = -\frac{1}{m}\left[\sum_{i=1}^{m \wedge y_i = 1} \frac{1}{\cancel{\sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)}}\left(\cancel{\sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)}\left(1 - \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)\right)\right) - \sum_{i=1}^{m \wedge y_i = 0} \frac{1}{\cancel{1 - \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)}}\left(\sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)\cancel{\left(1 - \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)\right)}\right)\right]$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{1}{m}\left[\sum_{i=1}^{m \wedge y_i = 0} \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right) - \sum_{i=1}^{m \wedge y_i = 1}\left(1 - \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)\right)\right]$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{1}{m}\left[\sum_{i=1}^{m}\left(1 - y_i\right)\sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right) - y_i\left(1 - \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)\right)\right]$$

$$\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{1}{m}\left[\sum_{i=1}^{m} \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right) \cancel{- y_i \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)} - y_i \cancel{+ y_i \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right)}\right]$$

$$\boxed{\frac{\partial}{\partial b} J\left(\mathbf{X}, \mathbf{y}, \mathbf{w}, b\right) = \frac{1}{m}\sum_{i=1}^{m} \sigma\left(\mathbf{w}\mathbf{x}_i^{\top} + b\right) - \mu\left(\mathbf{y}\right)}$$