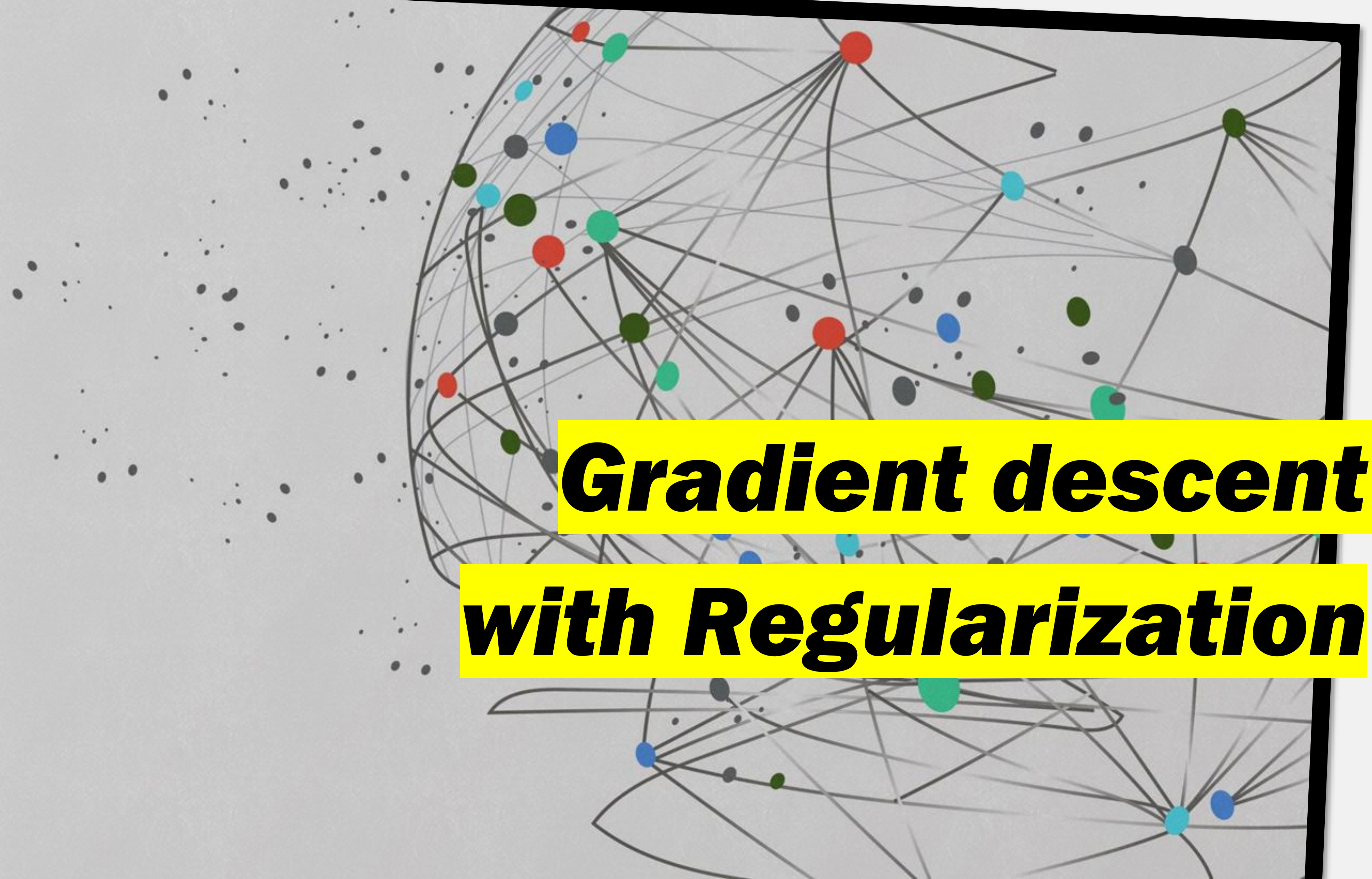
# Gradient descent with Regularization



$$\min_{\vec{w},b} J(\vec{w},b) = \min_{\vec{w},b} \left\{ \frac{1}{2m} \sum_{i=1}^{m} \left[ f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right]^2 + \frac{\lambda}{2m} \sum_{j=1}^{n} w_j^2 \right\}$$

Gradient descent

repeat {

$$w_j = w_j - \alpha \boxed{\frac{\partial}{\partial w_j} J(\vec{w},b)} \rightarrow \frac{1}{m} \sum_{i=1}^{m} \left( f_{\vec{w},b}(x^{(i)}) - y^{(i)} \right) x_j^{(i)} + \frac{\lambda}{m} w_j$$

$$b = b - \alpha \boxed{\frac{\partial}{\partial b} J(\vec{w},b)} \rightarrow \frac{1}{m} \sum_{i=1}^{m} \left[ f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right]$$

$$w_j = \underbrace{\left(1 - \frac{\alpha\lambda}{m}\right)} w_j - \underbrace{\frac{\alpha}{m}\sum_{i=1}^{m}\left[f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)}\right] x_j^{(i)}}_{\text{Usual update}}$$

$$\left.\begin{array}{l}\alpha = 0.01 \\ \lambda = 1 \\ m = 100\end{array}\right\} \rightarrow 0.9999$$

$$\frac{\partial}{\partial w_j} J(\vec{w}, b) = \frac{\partial}{\partial w_j}\left[\frac{1}{2m}\sum_{i=1}^{m}\left[f(\vec{x}^{(i)}) - y^{(i)}\right]^2 + \frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2\right]$$

$$\boxed{f(\vec{x}^{(i)}) = \vec{w}\,\vec{x}^{(i)} + b}$$

$$= \frac{1}{2m}\sum_{i=1}^{m} 2\cdot\left[\vec{w}\,\vec{x}^{(i)} + b - y^{(i)}\right]\cdot \vec{x}_j^{(i)} + \frac{2\lambda}{2m} w_j$$

$$= \frac{1}{m}\left[\sum_{i=1}^{m}\left[f(\vec{x}^{(i)}) - y^{(i)}\right]\cdot \vec{x}_j^{(i)} + \lambda w_j\right]$$

# Types of Regularization

- Ridge Regression (L2 Regularization)

**Effect:** Shrinks coefficients smoothly, making the model more robust to noise.
**Use Case:** When all features are relevant, but we want to prevent large coefficients.

$$\mathrm{Loss} = \sum (y - \hat{y})^2 + \lambda \sum w^2$$

- Lasso Regression (L1 Regularization)

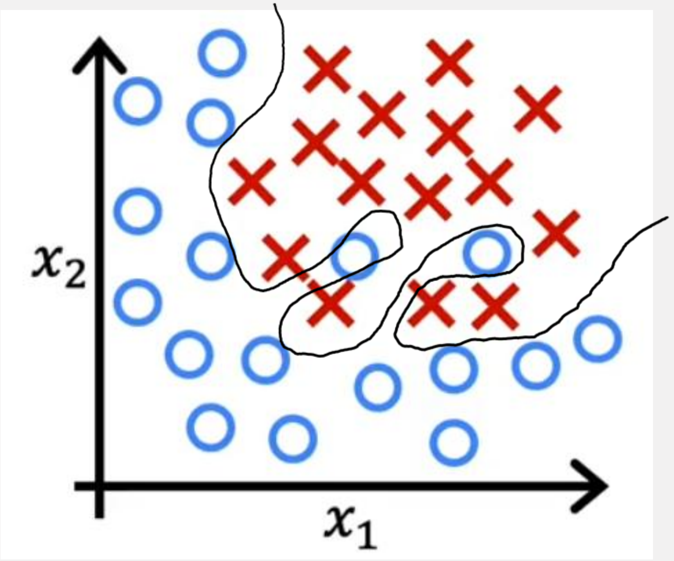
**Effect:** Shrinks some coefficients to exactly zero, making it useful for feature selection.
**Use Case:** When we suspect that some features are unnecessary.

$$\mathrm{Loss} = \sum (y - \hat{y})^2 + \lambda \sum |w|$$

# Dataset

| x | y |
| --- | --- |
| 1 | 2 |
| 2 | 4 |
| 3 | 6 |
| 4 | 8 |
| 5 | 10 |
| 6 | 13 |
| 7 | 15 |
| 8 | 20 |
| 9 | 25 |
| 10 | 30 |

# Logistic Regression




$$J(\vec{w}, b) = -\frac{1}{m} \sum_{i=1}^{m} \left[ y^{(i)} \log\left(f_{\vec{w},b}(\vec{x}^{(i)})\right) + (1 - y^{(i)}) \log\left(1 - f_{\vec{w},b}(\vec{x}^{(i)})\right) \right]$$

$$+ \frac{\lambda}{2m} \sum_{j=1}^{n} w_j^2$$

# Gradient Descent <u>L.R.</u>

$$w_j = w_j - \alpha \frac{\partial}{\partial w_j} J(\vec{w}, b) \rightarrow \frac{1}{m} \sum_{i=1}^{m} \left[ \left[ f_{\vec{w},b}(x^{(i)}) - y^{(i)} \right] x_j^{(i)} \right]$$

$$+ \frac{d}{m} w_j$$

$$b = b - \alpha \frac{\partial}{\partial b} J(\vec{w}, b) \rightarrow \frac{1}{m} \sum_{i=1}^{m} \left[ f_{\vec{w},b}(\vec{x}^{(i)}) - y^{(i)} \right]$$