ENGR 3321: Introduction to Deep Learning for Robotics

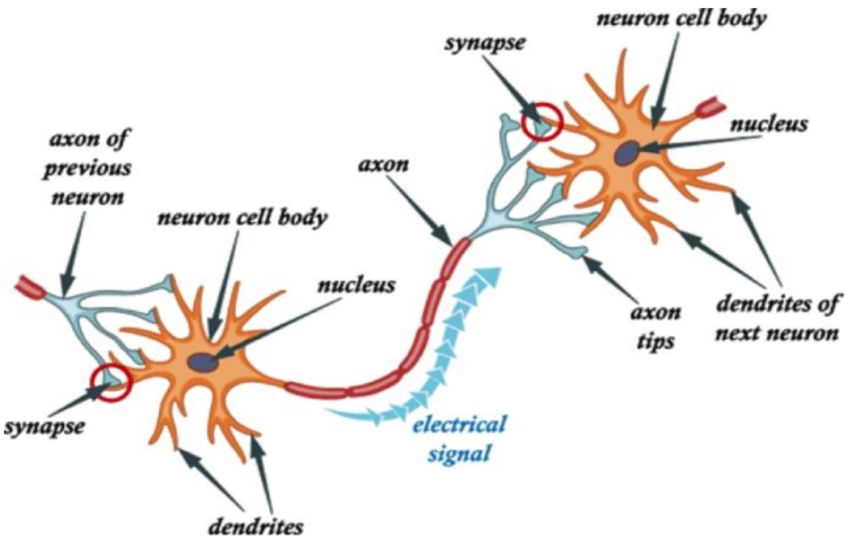
Neural Network 101: SISO Linear Function

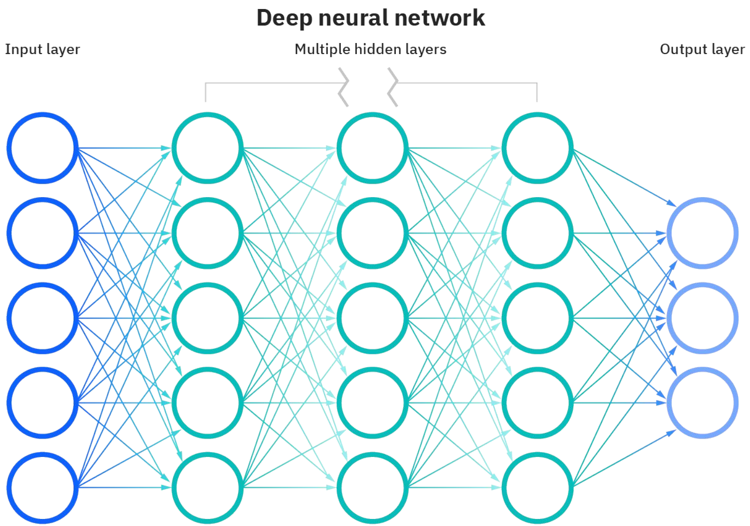
09/18/2023

Outline

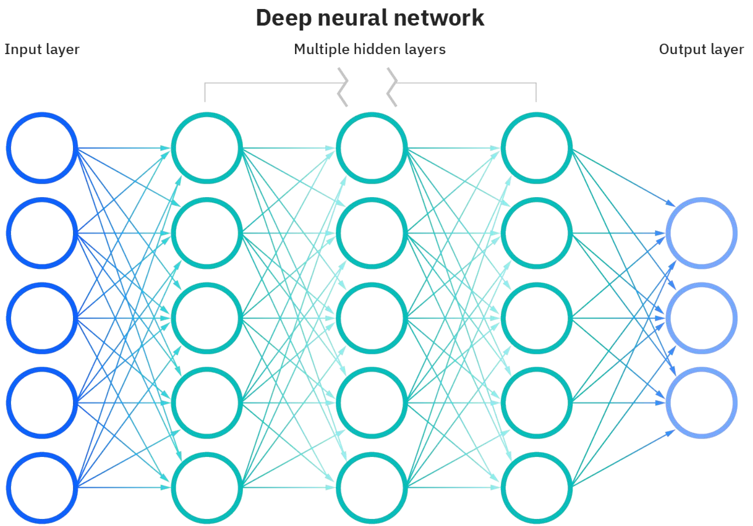
- Forward Pass
- Loss Function
- Gradient Descent
- Linear Model

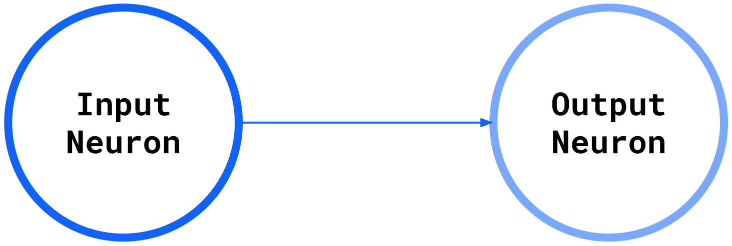
Neural Network



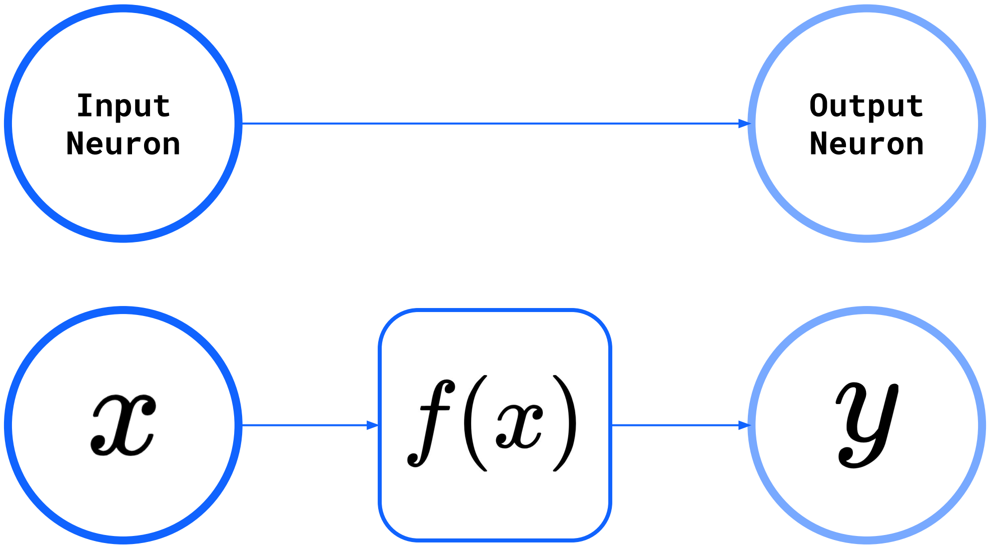


Simplified Neural Network





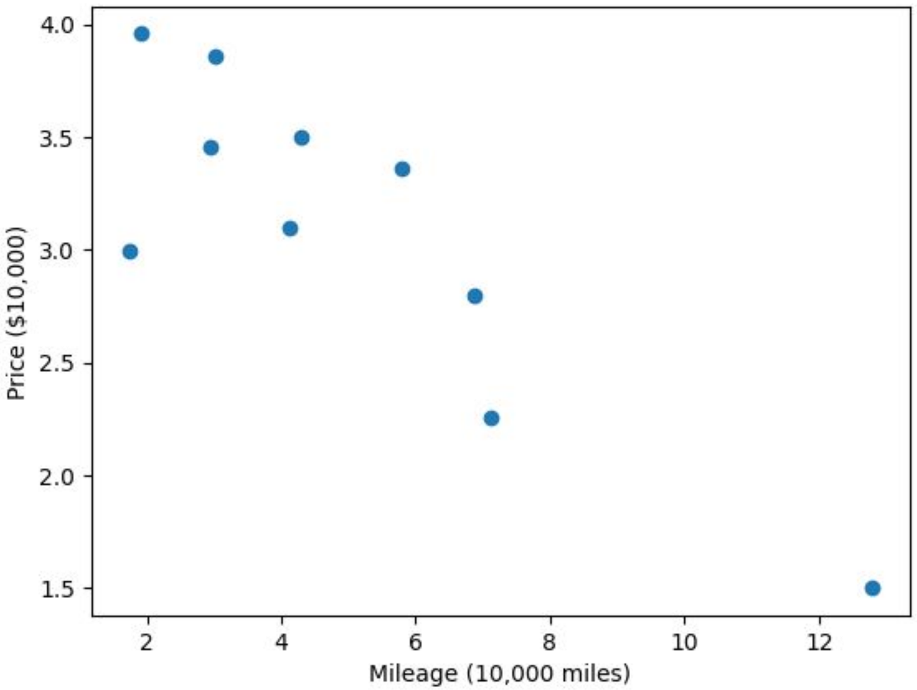
Neural Network == Function



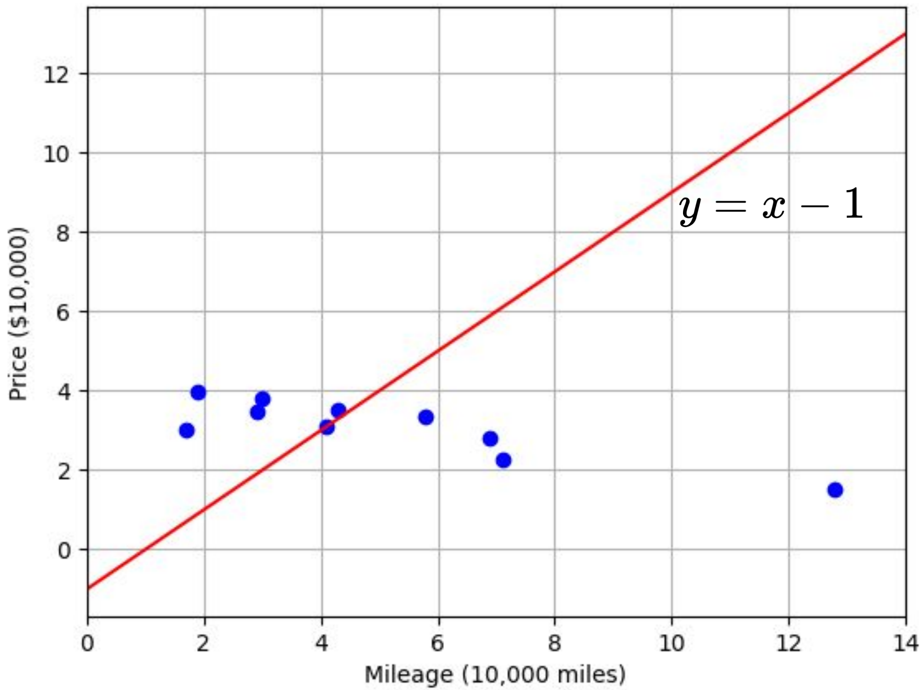
$$y = f(x)$$

Example: Predict Used Car Price

Mileage (10,000 miles)	Selling Price ($10,000)
5.7923	3.359
7.1229	2.259
1.9160	3.959
4.1124	3.099
12.8000	1.5
6.8696	2.799
2.9499	3.459
4.3000	3.5
1.7302	2.999
3.0237	3.859

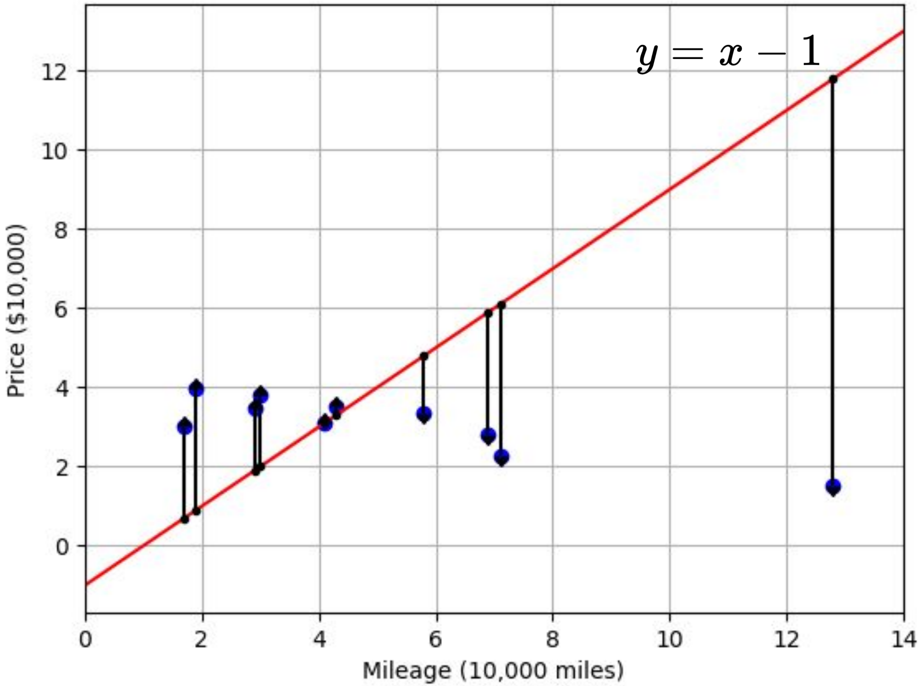


Initial Guess



Model Performance

Dataset: $\left\{\left(x^{(1)}, y^{(1)}\right), \left(x^{(2)}, y^{(2)}\right), \ldots, \left(x^{(M)}, y^{(M)}\right)\right\}$



$$\mathcal{L}(\hat{y}, y) = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{2}(\hat{y}_i - y_i)^2$$

Gradient

$$\nabla \mathcal{L}(w,b) = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial w}(w,b) \\ \frac{\partial \mathcal{L}}{\partial b}(w,b) \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathcal{L}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w} \\ \frac{\partial \mathcal{L}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial b} \end{bmatrix} = \begin{bmatrix} \frac{1}{M} \sum_{i=1}^{M} (\hat{y}_i - y_i) x_i \\ \frac{1}{M} \sum_{i=1}^{M} (\hat{y}_i - y_i) \end{bmatrix}$$

Chain Rule

Gradient Descent

Given dataset: $\{(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)\}$

Initialize $w \, and \, b$

Repeat until converge {

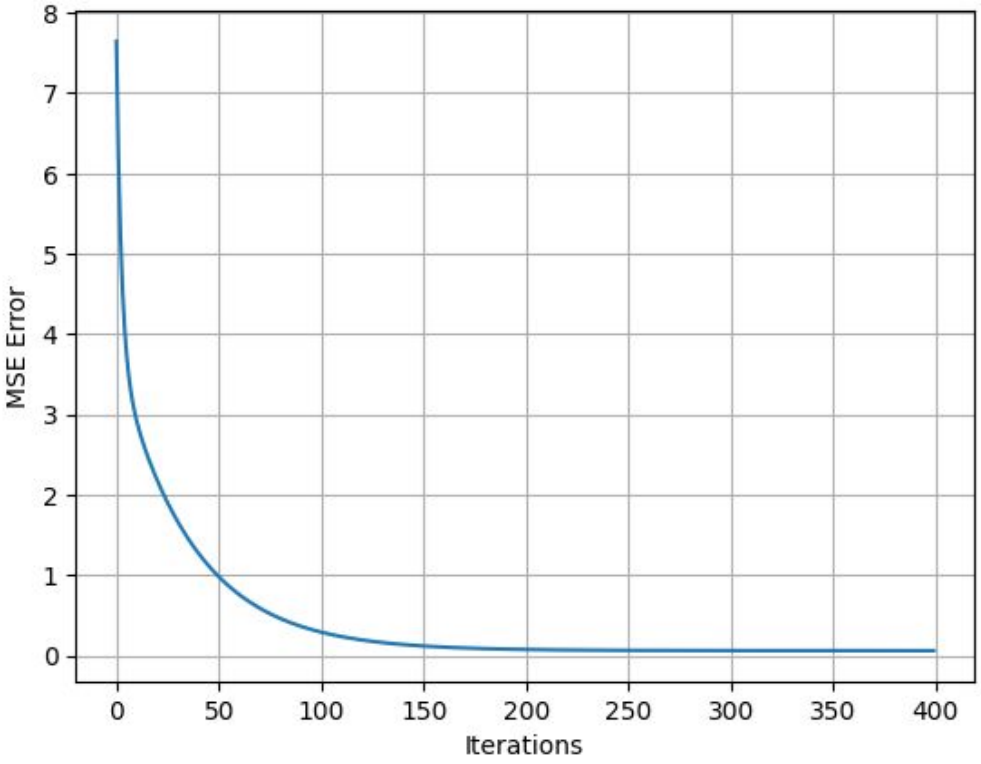
$$w := w - \alpha \frac{\partial \mathcal{L}}{\partial w}$$

$$b := b - \alpha \frac{\partial \mathcal{L}}{\partial w}$$

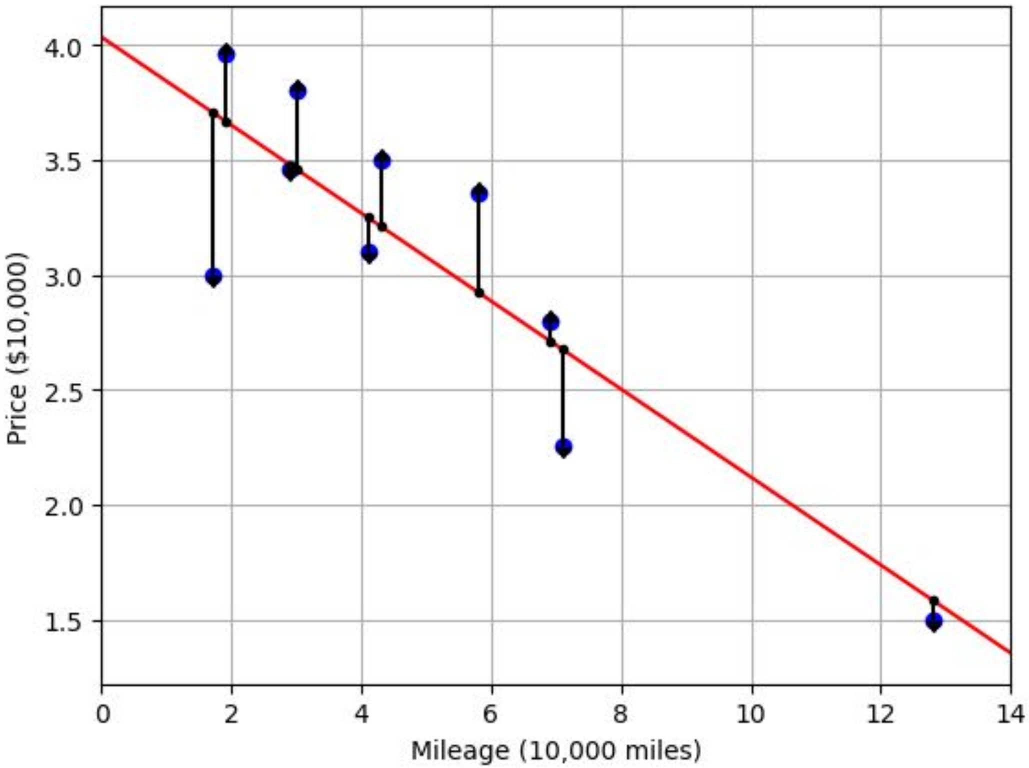
}

where α is learning rate

Loss Decrease



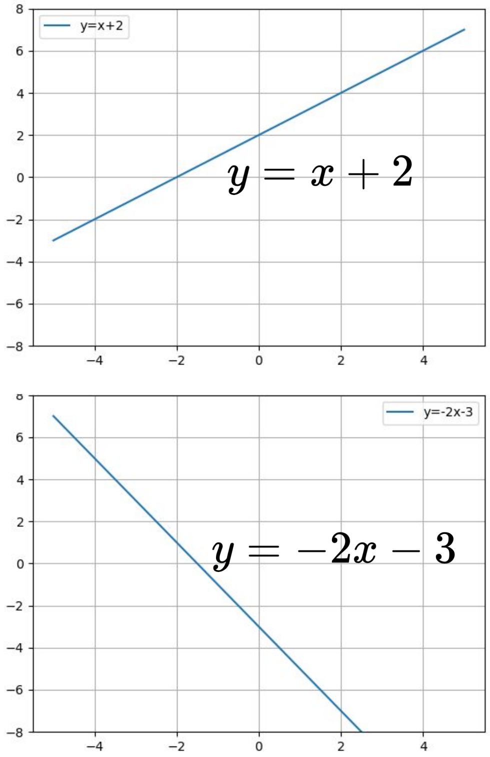
Trained Model



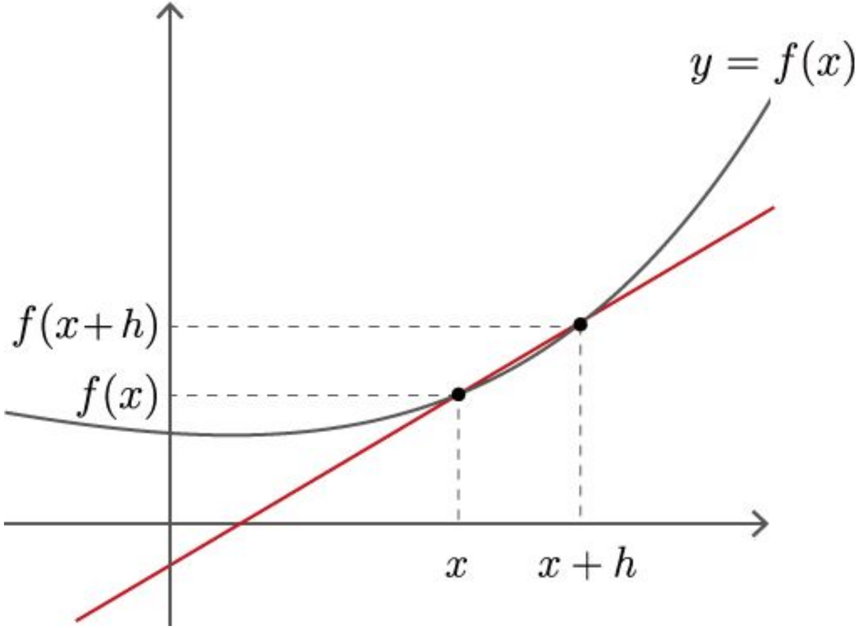
Linear Model Review

$$y = f(x)$$

$$y = wx + b$$

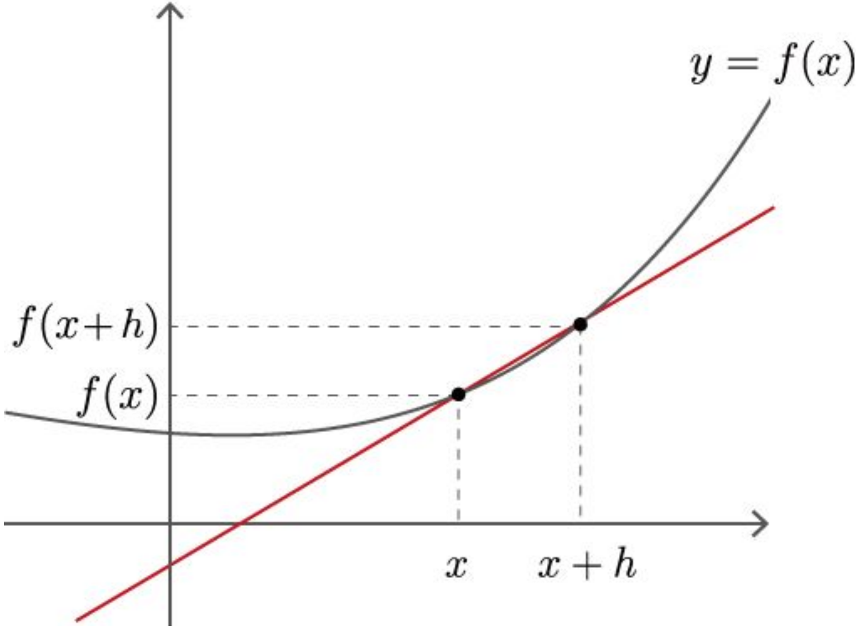


Derivative Review



the derivative of a function at a given point gives us the rate of change or slope of the tangent line to the function at that point.

Analytic Derivative vs. Numerical Derivative



Analytic derivative: $f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$

- Fast
- Accurate
- Error-Prone

Numerical derivative:

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$

- Slow
- Approximate
- Easy to code

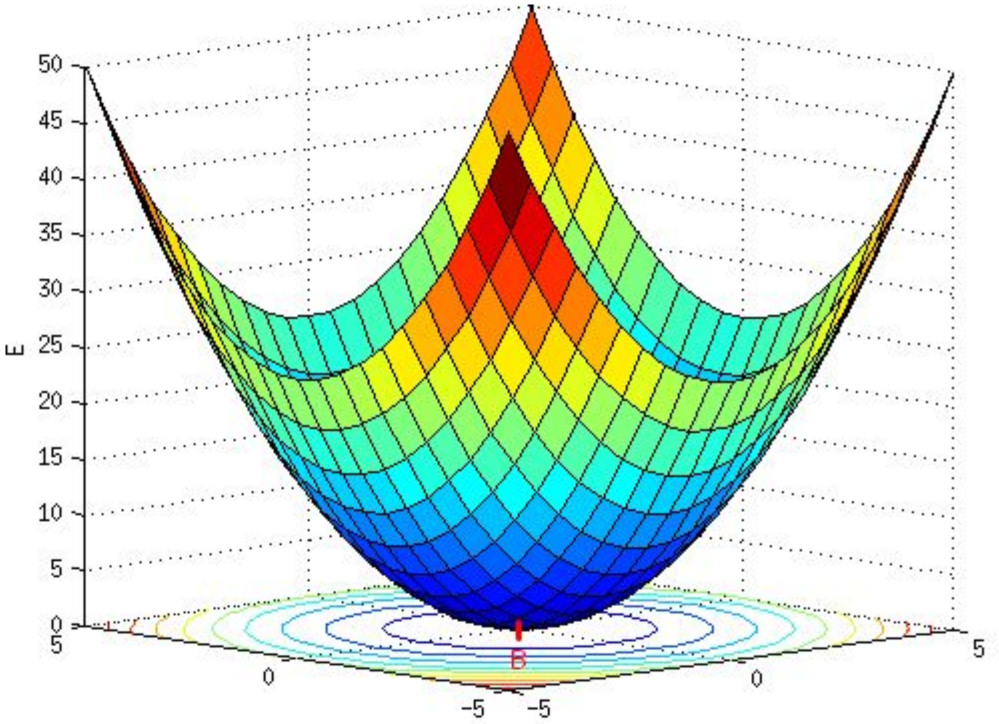
Chain Rule Review

The Chain Rule

If $y = f(u)$, where $u = g(x)$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

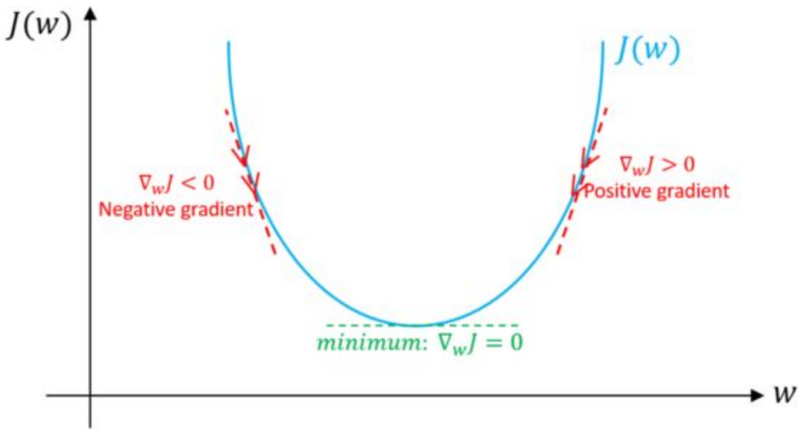
Gradient Review

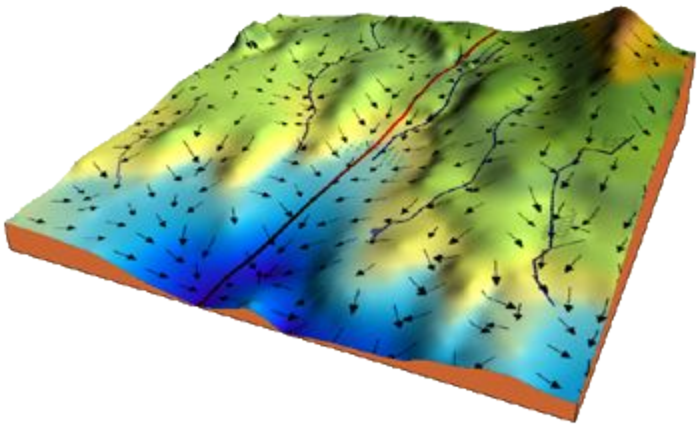


The gradient stores all the partial derivative information of a multivariable function.

$$\nabla f(x_1, \ldots, x_n) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x_1, \ldots, x_n) \\ . \\ . \\ . \\ \frac{\partial f}{\partial n}(x_1, \ldots, x_n) \end{bmatrix}$$

Gradient Descent Concept





Find w and b that minimize $\mathcal{L}(w, b)$