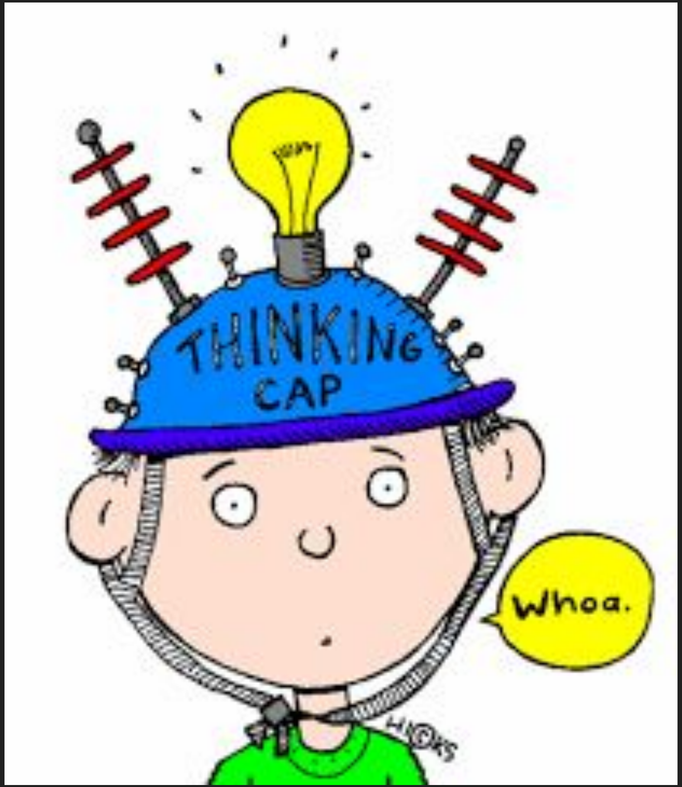
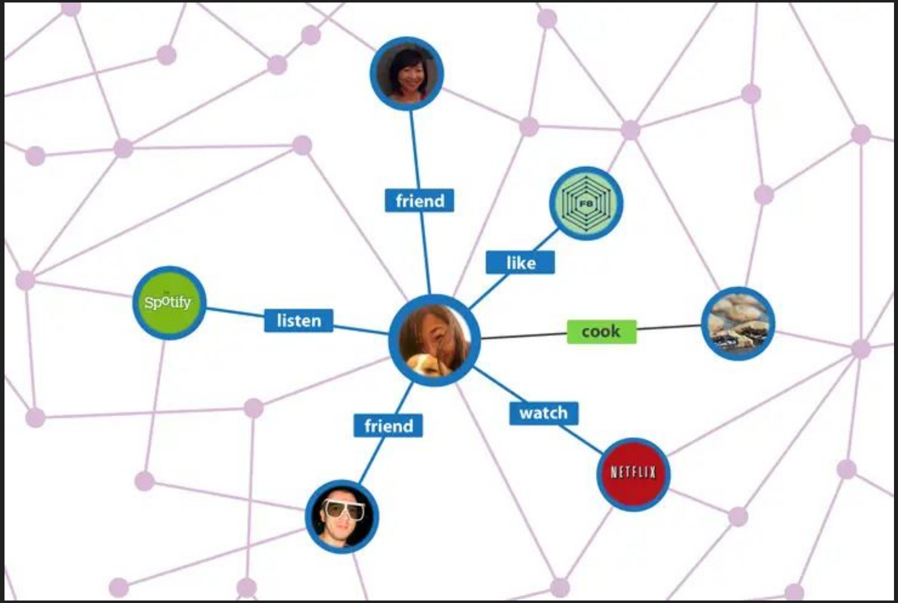
L03

Computation Graph.
Matrix Operation in Pytorch.
Linear Regression. Again.

Any Questions on HW?

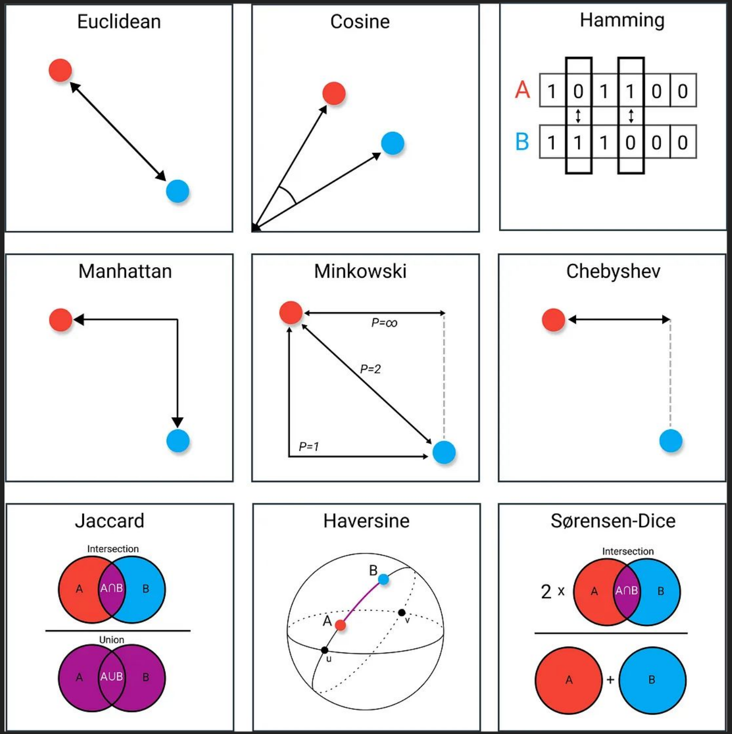


Graph


friend
like
listen
cook
watch
friend
Spotify
NETFLIX


(x_2,y_2)
(x_1,y_1)

Nodes and Edges. Distances.

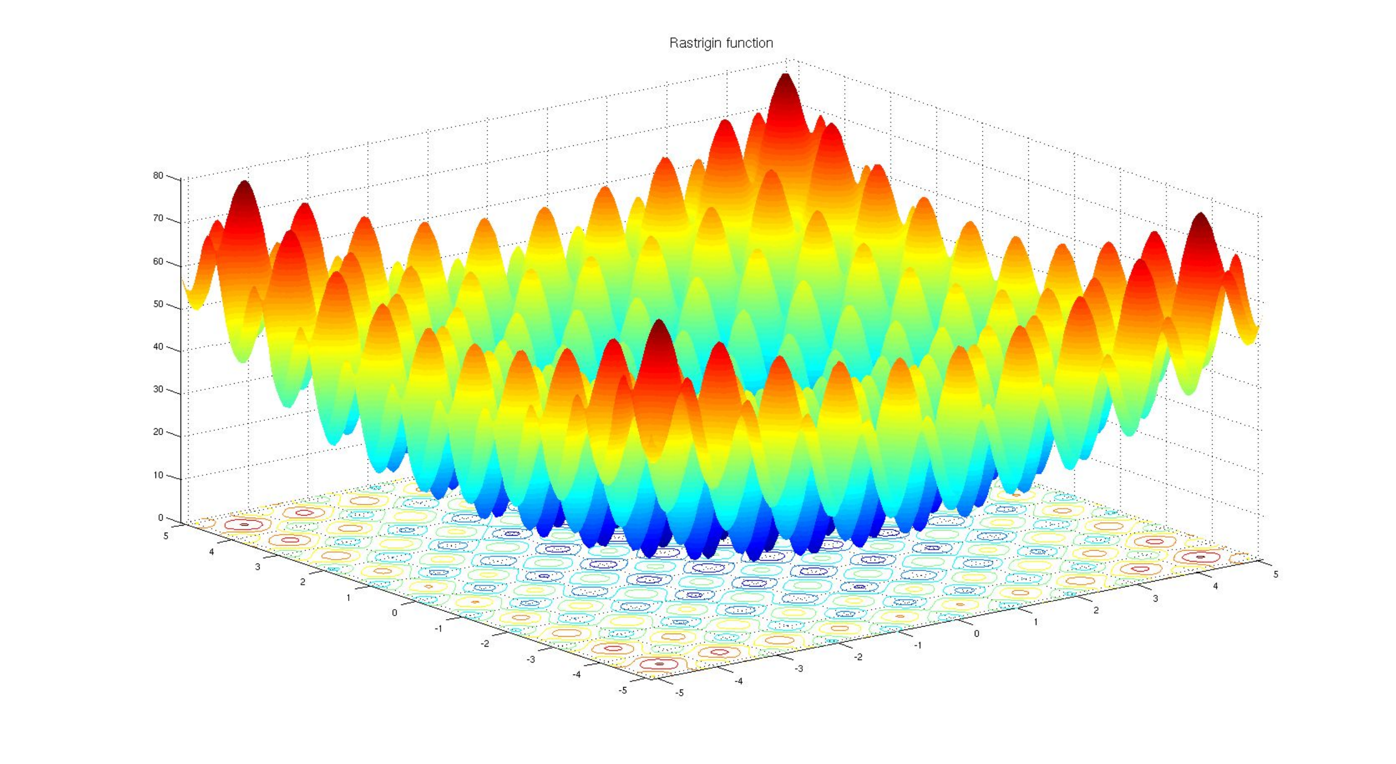


Ref: https://towardsdatascience.com/9-distance-measures-in-data-science-918109d069fa

Talk about Minimization

1. What task does ML solve?
2. How does it solve?
3. How minimization task is solved?
4. What is derivative?
5. How are derivative and minima are connected?
6. What is the difference between gradient and derivative?
7. Why do we need to know how to find function minima?

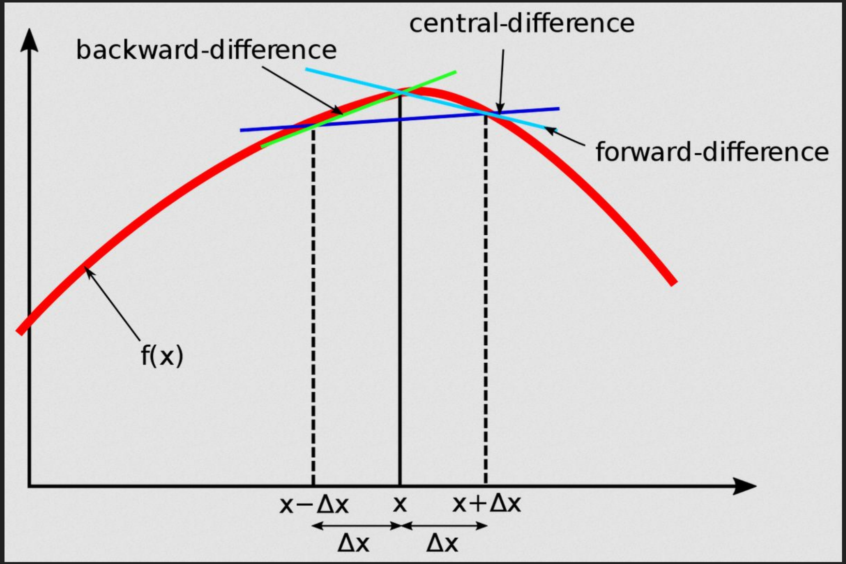
Test Functions for Optimization



Ref: https://en.wikipedia.org/wiki/Test_functions_for_optimization

Derivative. How to Find It?

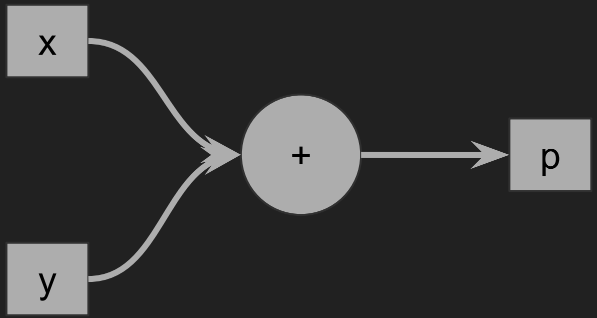
- Analytic differentiation
- Numerical differentiation



Calculation graph is about programming of analytic differentiation.

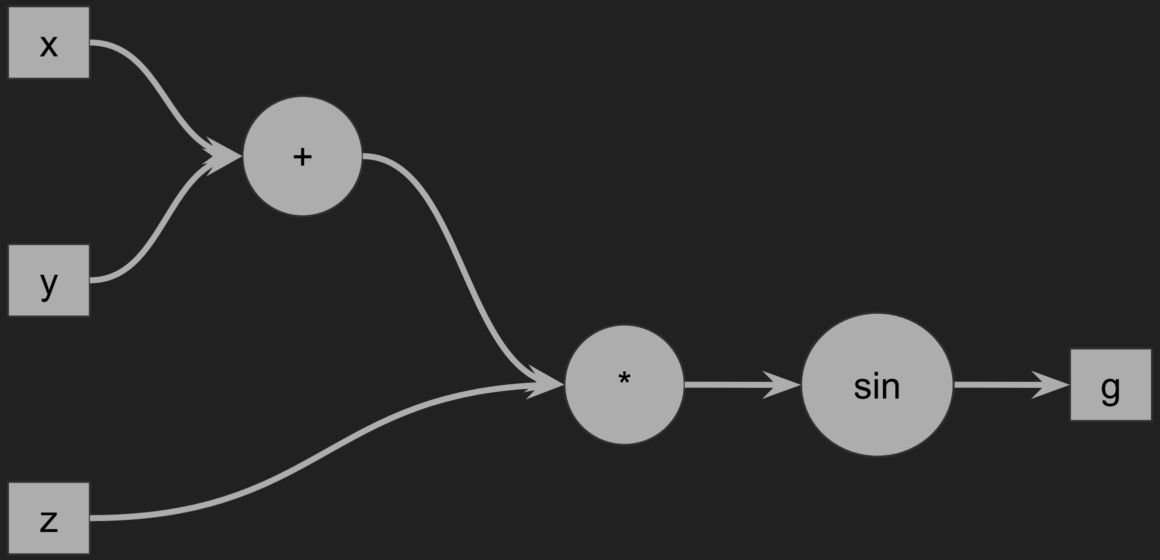
Computation Graph. Example

$$p = x + y$$



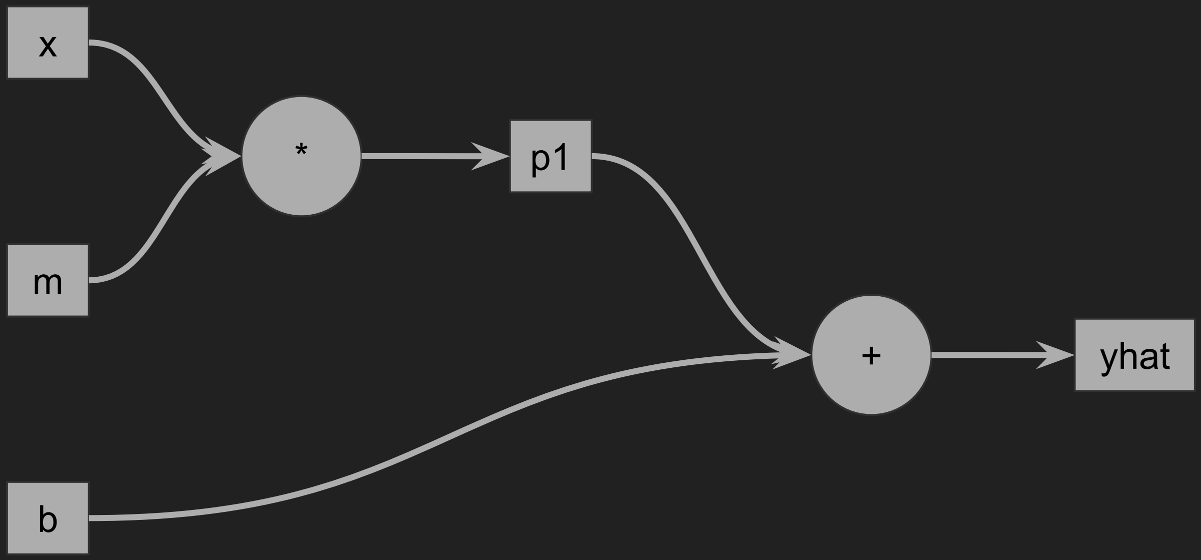
Computation Graph. Example

$$g = \sin((x + y) * z)$$



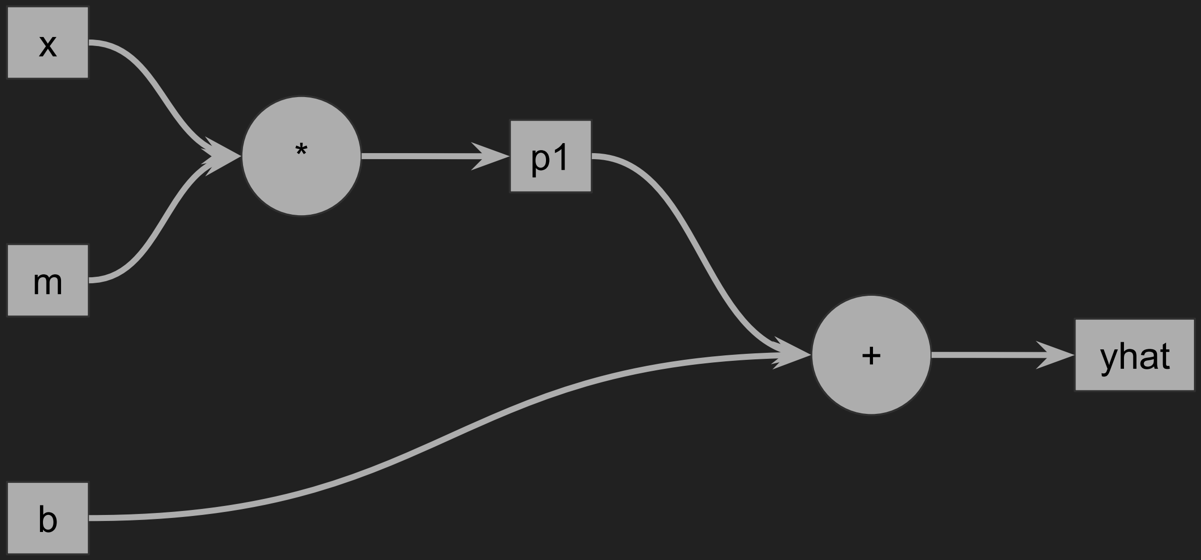
Computation Graph. Linear Regression. Forward Pass.

*yhat = m*x + b*



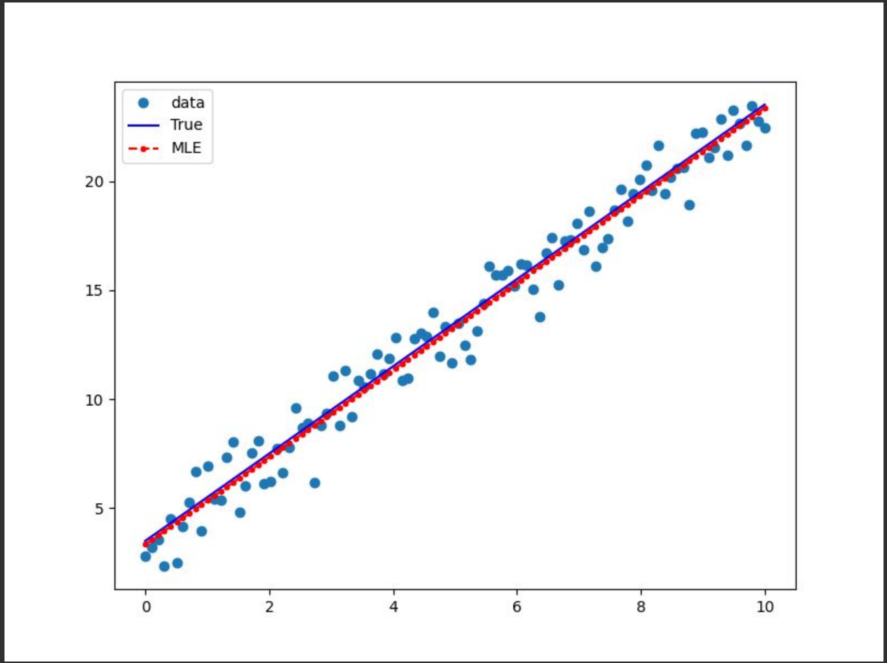
yhat = ? for x = 1, m = 2, b = 1

*yhat = m*x + b*



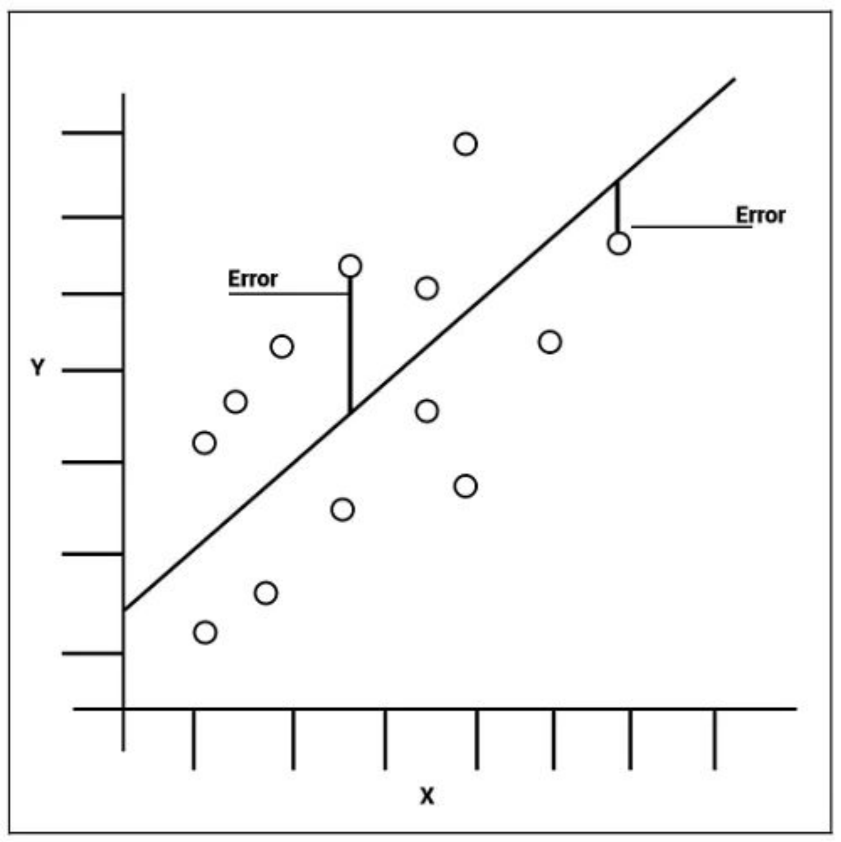
How to find params *m* and *b*?

Minimize loss function

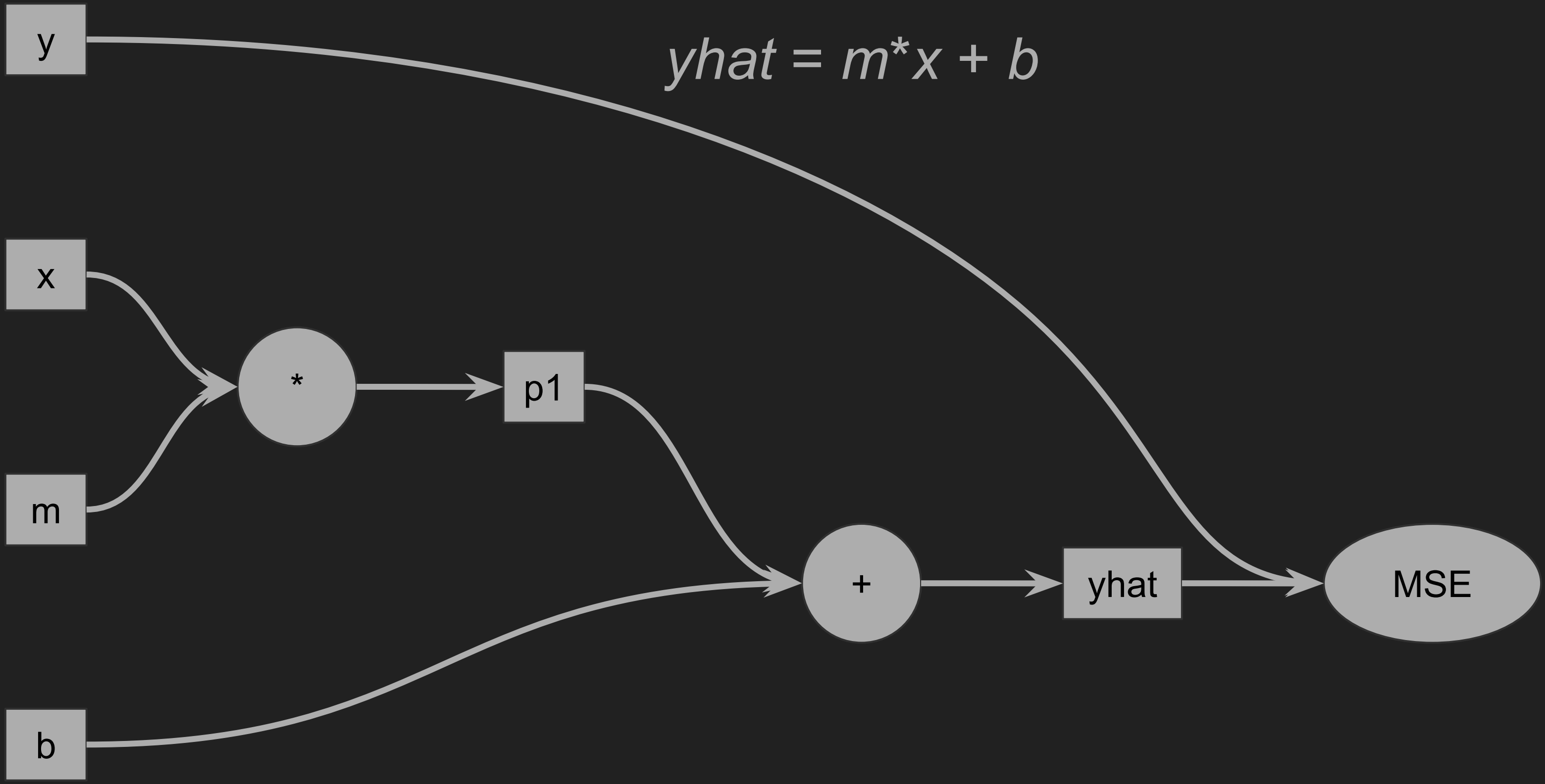


Loss function

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad \text{where} \quad \hat{y}_i = mx_i + b$$

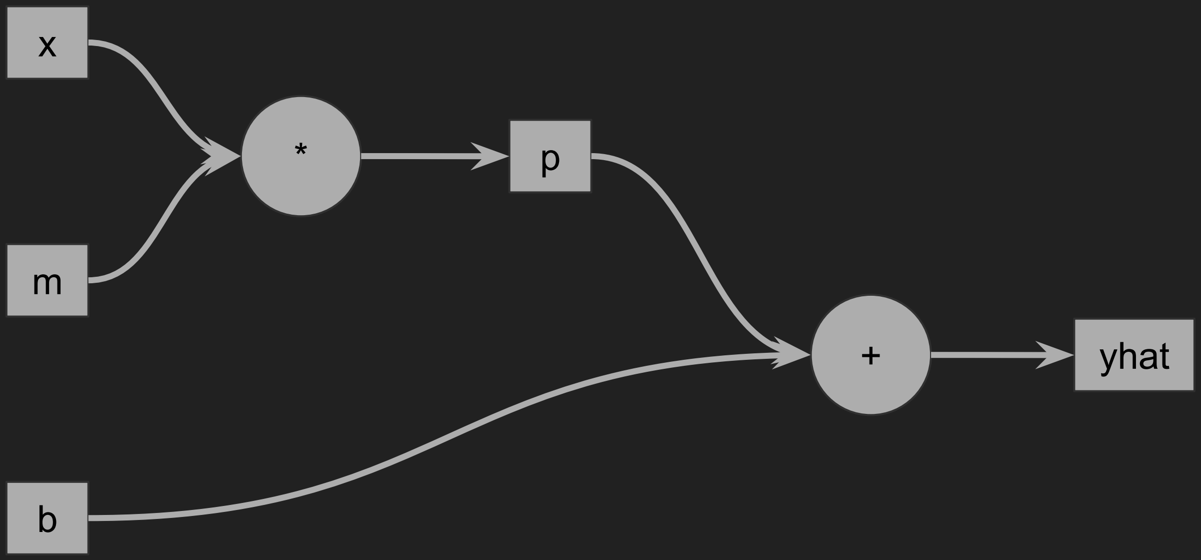


Loss Function in Computation Graph



Derivatives with Graphs

Derivatives



$$p = m * x$$

$$\widehat{y} = p + b$$

$$\frac{\mathrm{d}\widehat{y}}{\mathrm{d}p} = 1$$

$$\frac{\mathrm{d}\widehat{y}}{\mathrm{d}b} = 1$$

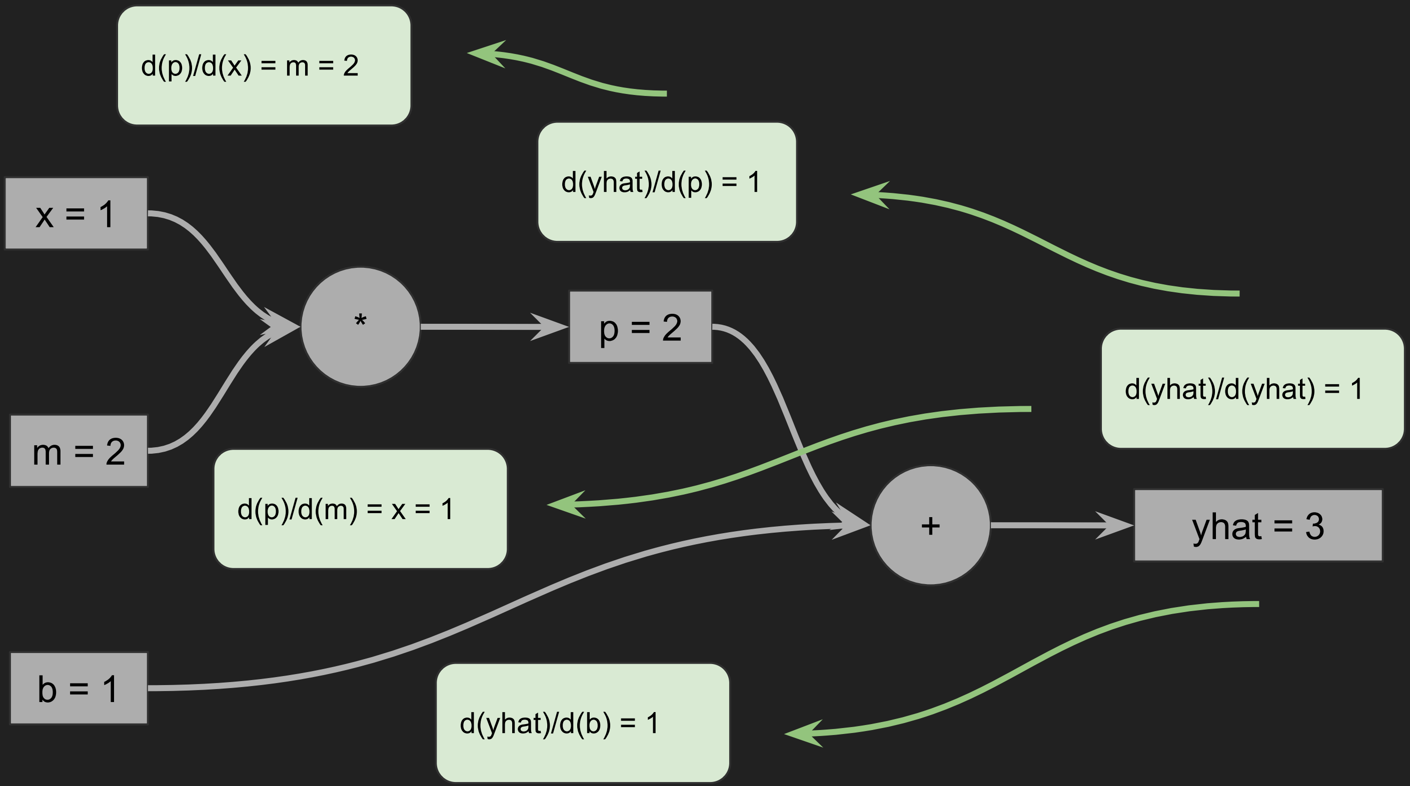
$$\frac{\mathrm{d}p}{\mathrm{d}m} = x$$

$$\frac{\mathrm{d}p}{\mathrm{d}x} = m$$

$$\frac{\mathrm{d}\widehat{y}}{\mathrm{d}m} = \frac{\mathrm{d}\widehat{y}}{\mathrm{d}p}\frac{\mathrm{d}p}{\mathrm{d}m}$$

$$\frac{\mathrm{d}\widehat{y}}{\mathrm{d}x} = \frac{\mathrm{d}\widehat{y}}{\mathrm{d}p}\frac{\mathrm{d}p}{\mathrm{d}x}$$

Derivatives



$$p = m * x$$

$$\hat{y} = p + b$$

$$\frac{d\hat{y}}{dp} = 1$$

$$\frac{d\hat{y}}{db} = 1$$

$$\frac{dp}{dm} = x$$

$$\frac{dp}{dx} = m$$

$$\frac{d\hat{y}}{dm} = \frac{d\hat{y}}{dp}\frac{dp}{dm}$$

$$\frac{d\hat{y}}{dx} = \frac{d\hat{y}}{dp}\frac{dp}{dx}$$

Backpropagation Example

L(a, b, c) = c(a + 2b)

d = 2b

e = a + d

L = c*e

$$L = ce \;:\; \frac{\partial L}{\partial e} = c, \frac{\partial L}{\partial c} = e$$

$$e = a + d \;:\; \frac{\partial e}{\partial a} = 1, \frac{\partial e}{\partial d} = 1$$

$$d = 2b \;:\; \frac{\partial d}{\partial b} = 2$$

$$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial e}\frac{\partial e}{\partial a}$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial e}\frac{\partial e}{\partial d}\frac{\partial d}{\partial b}$$

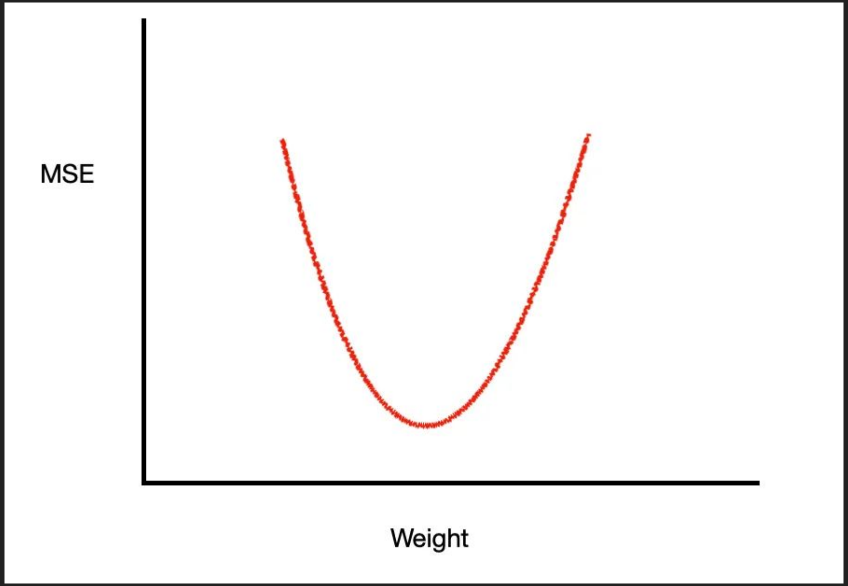
Loss Function Backpropagation

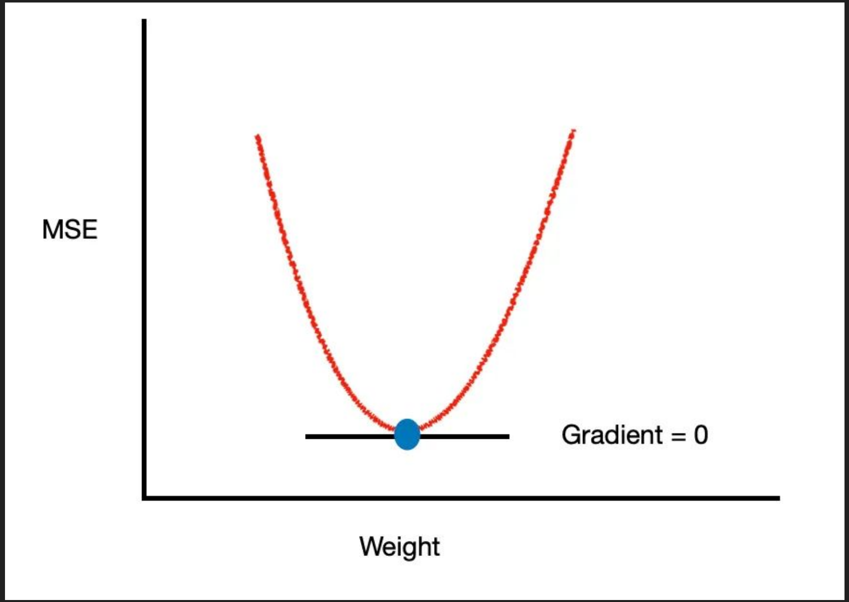
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad \text{where} \quad \hat{y}_i = mx_i + b$$

$$\frac{\partial f}{\partial m} = \frac{1}{n}\sum_{i=1}^{n} -2x_i(y_i - (mx_i + b))$$

$$\frac{\partial f}{\partial b} = \frac{1}{n}\sum_{i=1}^{n} -2(y_i - (mx_i + b))$$

Gradient Descent

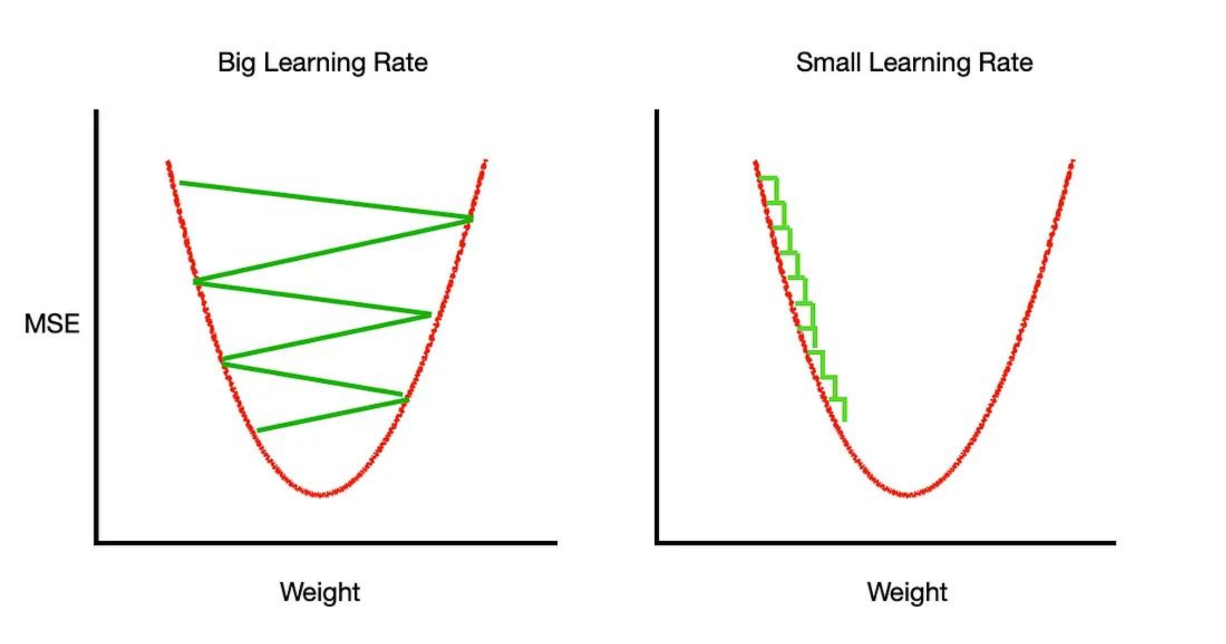




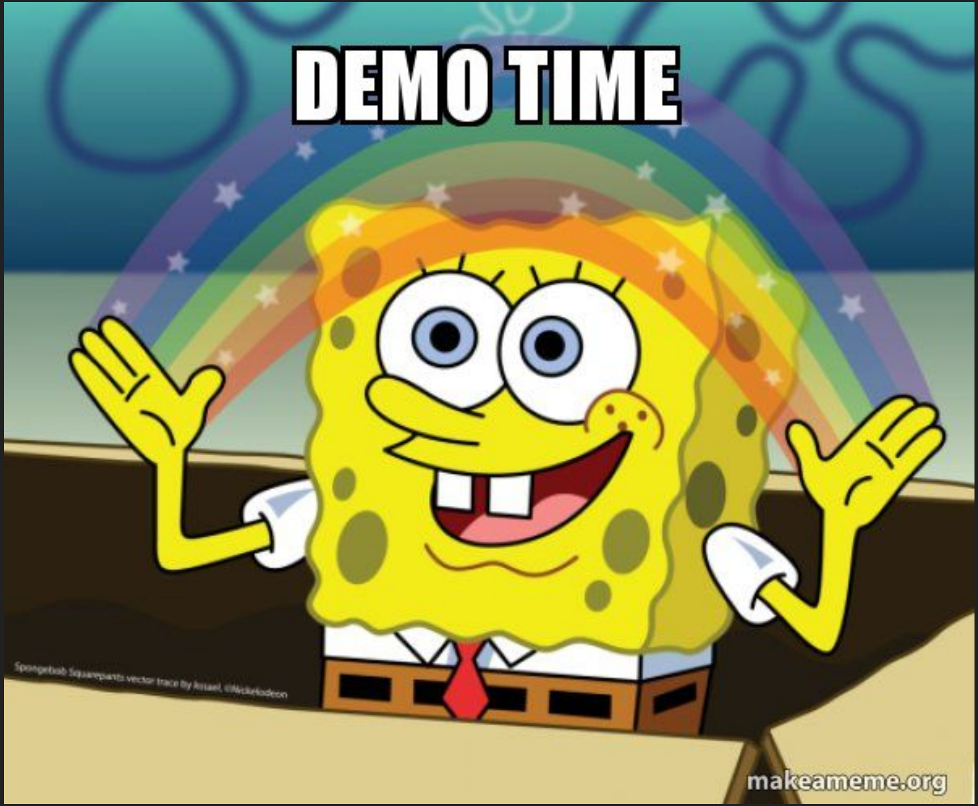
https://www.geogebra.org/m/xC6zq7Zv

Gradient Descent Learning Rate

$$w_{i+1}^{j} = w_{i}^{j} - l_r \frac{dL}{dw^j}$$



Gradient Descent Demo



HW

- Take analytical derivative of sigmoid function
- Experiments with demo code (gradient descent)
    - Vary learning rate
    - Vary epochs
    - Plot MSE over training (over epochs for specific learning rate)
- Make one forward and backward steps for
    - $L = (2a + b)(c - d)$,

      a, b, c, d are arbitrary numbers