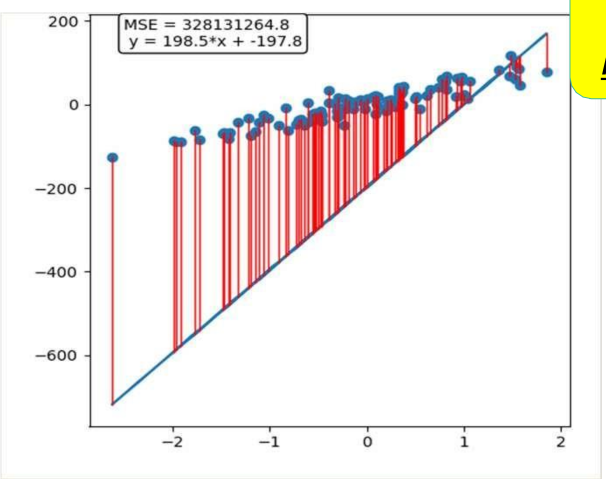
Data Science for Engineers

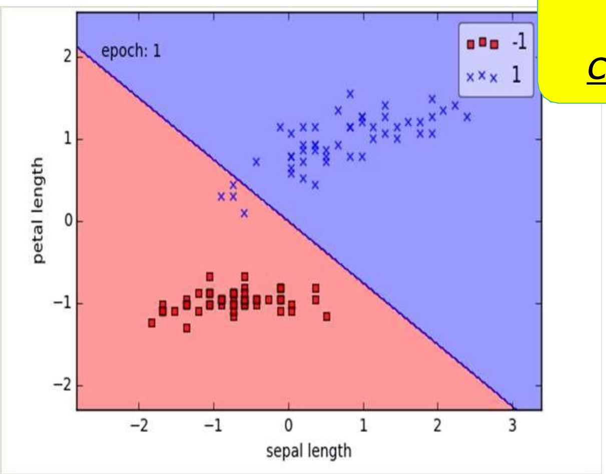
WEEK 4 THEORY

Optimization

- A optimization problem consists of maximizing or minimizing a real function by systematically choosing input value from within as allowed set and computing the value of the function (source: Wikipedia)
- Use of specific method to determine the ***'best'*** solution to the problem



Best line to *represent data*



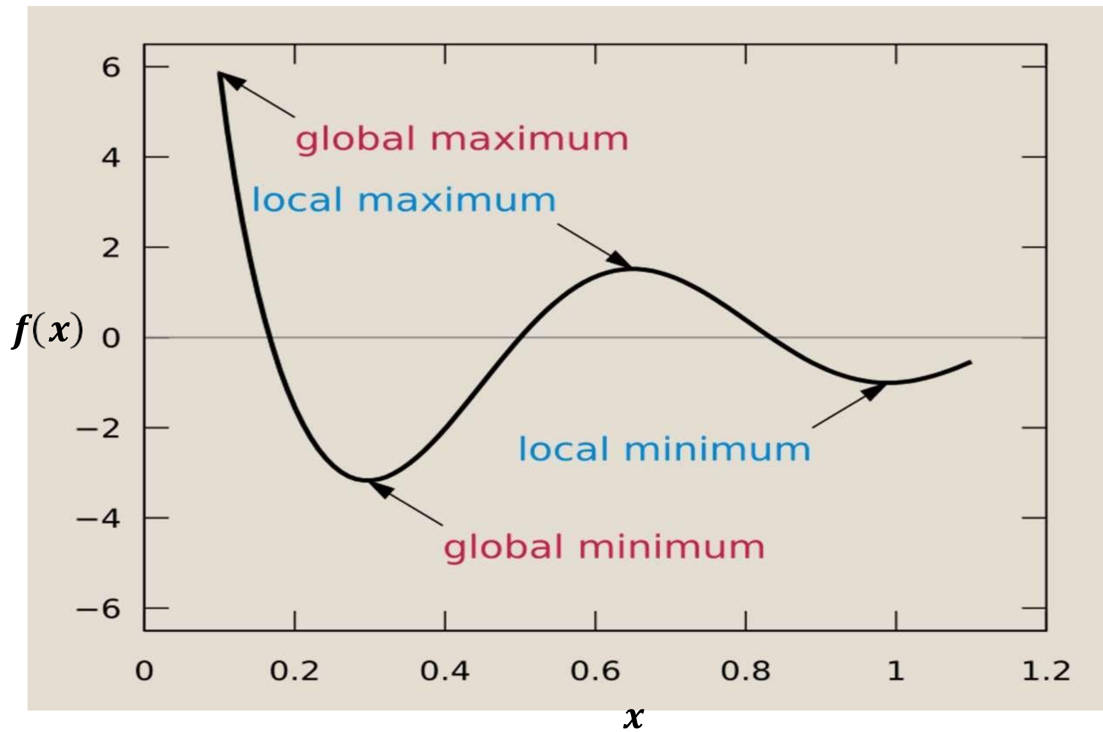
Best line to *classify data*

2

Univariate Optimization problem

- Objective function
- Decision variable
- Constraints

$$\min_{x} f(x)$$
$$x \in R$$



Univariate optimization conditions

$$\min_{x} f(x)$$
$$x \in R$$

- Necessary condition for x to be minimizer

$$f'(x) = 0$$

- Sufficient condition

$$f''(x) > 0$$

Que: The minima /maxima of f(x) exist when

a) $f'(x) > 0$
b) $f'(x) = 0$
c) $f'(x) < 0$

Que: The maxima of f(x) exist when

a) $f''(x) > 0$
b) $f''(x) = 0$
c) $f''(x) < 0$

5

Q: $f(x) = \frac{x^4}{4} - \frac{x^3}{3} - x^2$

a) Find the stationary points for the following points

b) Find the values of x at which minima exist and its value

Q: Of all rectangles of area 100, which has the smallest perimeter?

- Let l be the length and b be breadth of the rectangle

$$Area = l \times b, \qquad Perimeter = 2l + 2b$$

- Here, the optimization function is perimeter and decision variable is length

$$f(l) = 2l + 2b = 2l + 2\left(\frac{A}{l}\right) = 2l + \frac{200}{l}$$

- Applying first order necessary condition $f'(x) = 0$

$$f'(l) = 2 - \frac{200}{l^2} = 0$$

- Solving above equation gives length = +10,-10
- As we know length is cannot be negative hence the length of the rectangle is 10.
- Calculating the breadth=10, the perimeter of the rectangle is ***P =2(10)+2(10)= 40***

7

Q. A manufacturer determines that the daily avg of producing q units is determined by the number of units produced per day which minimize the avg cost ?

$$C(q) = 0.0001q^2 - 0.08q + 65 + (5000/q)$$

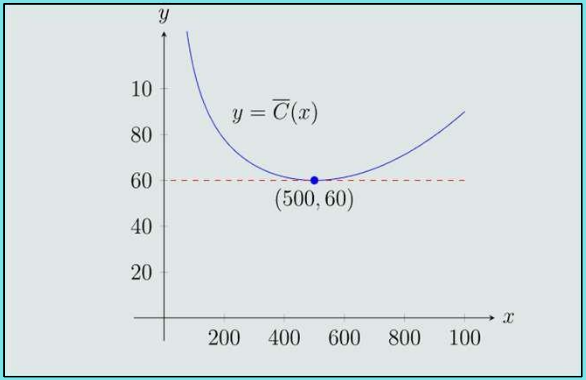
- Here, the optimization function is C(q) and q is the decision variable

$$C(q) = 0.0001q^2 - 0.08q + 65 + (\frac{5000}{q})$$

- Applying first order necessary condition $C'(q) = 0$

$$C'(q) = 0.0002q - 0.08 - (\frac{5000}{q^2})$$

- Solving above equation, the critical value **q= 500**



- Checking the second order necessity condition $C''(q) > 0$

$$C''(q) = 0.0002 + (\frac{10000}{q^3})$$

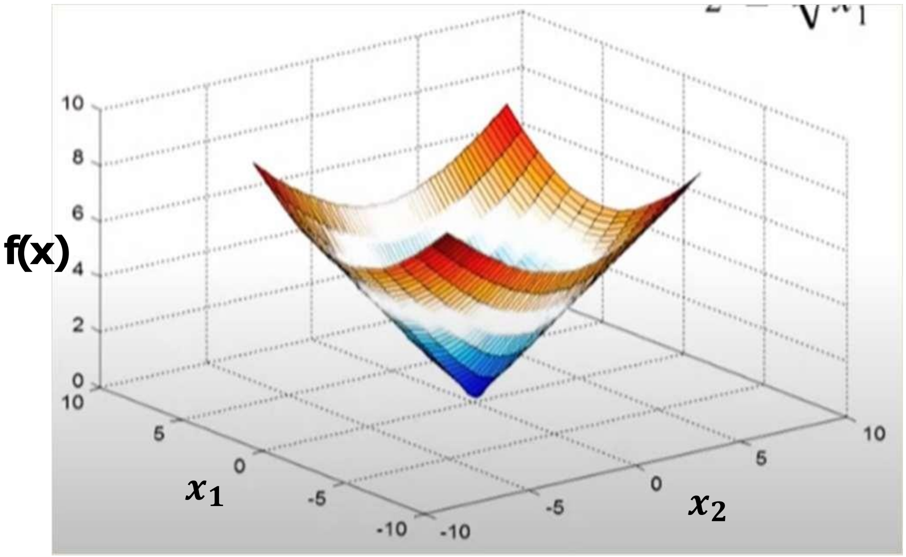
- Putting critical value into the equation give positive value $C''(q) > 0$. Hence the manufacturing cost can be minimize by producing **q = 500 units**

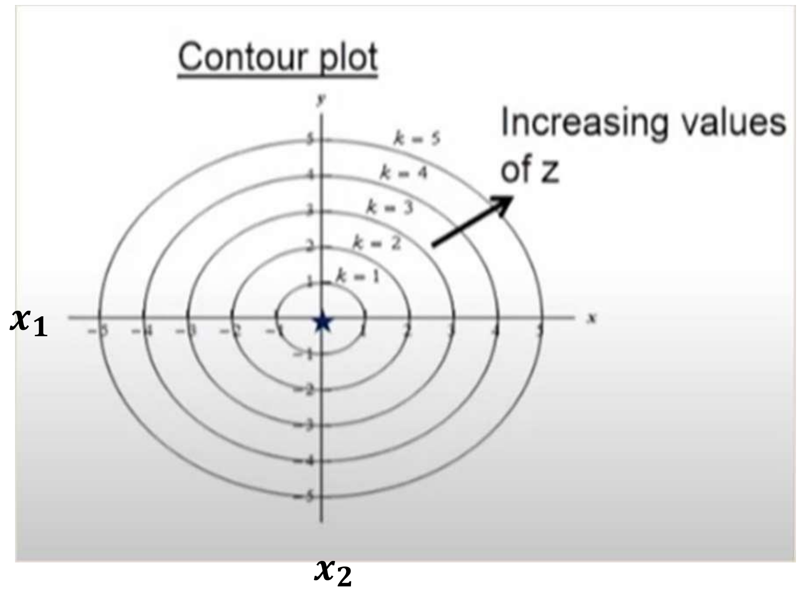
8

Multivariate optimization

$$z = \mathrm{f}(x_1, x_2, .., x_n)$$

- Let $z = \sqrt{{x_1}^2 + {x_2}^2}$





Multivariate optimization condition

$$z = \mathrm{f}(x_1, x_2, .., x_\mathrm{n})$$

Necessary condition for x to be minimizer : $\boldsymbol{\nabla f(x\ )} = \mathbf{0}$

$$\nabla f(x^*) = Gradient = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \dots \\ \dots \\ \partial f/\partial x_n \end{bmatrix}$$

- Sufficient condition : $\boldsymbol{\nabla^2 f(x^*)}$ **has to positive definite**

$$\nabla^2 f(x^*) = Hessian = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 & \dots & \partial^2 f/\partial x_1 \partial x_n \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 & \dots & \partial^2 f/\partial x_2 \partial x_n \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \partial^2 f/\partial x_n \partial x_1 & \partial^2 f/\partial x_n \partial x_2 & \dots & \partial^2 f/\partial x_n^2 \end{bmatrix}$$

Que: Find the 1st and 2nd order necessary conditions for the function and tell whether minima exists or not?

- The function is $f(x_1, x_2) = x_1 + 2x_2 + 4{x_1}^2 - x_1x_2 + 2{x_2}^2$
- Applying first order necessary condition $\nabla f = 0$

$$\nabla f = \begin{bmatrix} \frac{6f}{6x_1} \\ \frac{6f}{6x_2} \end{bmatrix} = \begin{bmatrix} 1 + 8x_1 - x_2 \\ 2 - x_1 + 4x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \end{bmatrix} = \begin{bmatrix} \mathbf{-0.19} \\ \mathbf{-0.54} \end{bmatrix}$$

- Checking the second order necessity condition i.e. $\boldsymbol{\nabla^2 f(x^*)}$ ***is positive definite***

$$\boldsymbol{\nabla^2 f} = \begin{vmatrix} \frac{\partial^2 f}{\partial {x_1}^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial {x_2}^2} \end{vmatrix} = \begin{vmatrix} 8 & -1 \\ -1 & 4 \end{vmatrix}$$

$|A - \lambda I| = 0$

$\frac{\partial f}{\partial x_1} = 1 + 8x_1 - x_2$

$\frac{\partial f}{\partial x_2} = 2 - x_1 + 4x_2$

$\frac{\partial^2 f}{\partial {x_1}^2} = 8$

$\frac{\partial^2 f}{\partial {x_2}^2} = 4$

$\frac{\partial^2 f}{\partial x_2 \partial x_1} = -1$

- The eigen values of the $\boldsymbol{\nabla^2}$f matrix ***are 3.76 and 8.23 >0***. Hence the matrix is positive definite and ***minima exist at the critical points***.

11

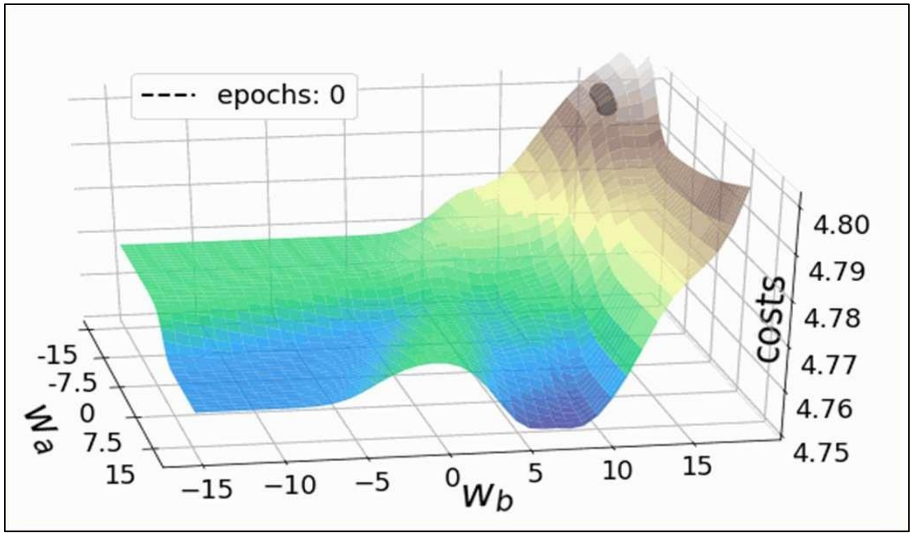
Gradient descent method

- Applicable to find the minima of the given function
- Applied in the backpropagation algorithm to find the parameter

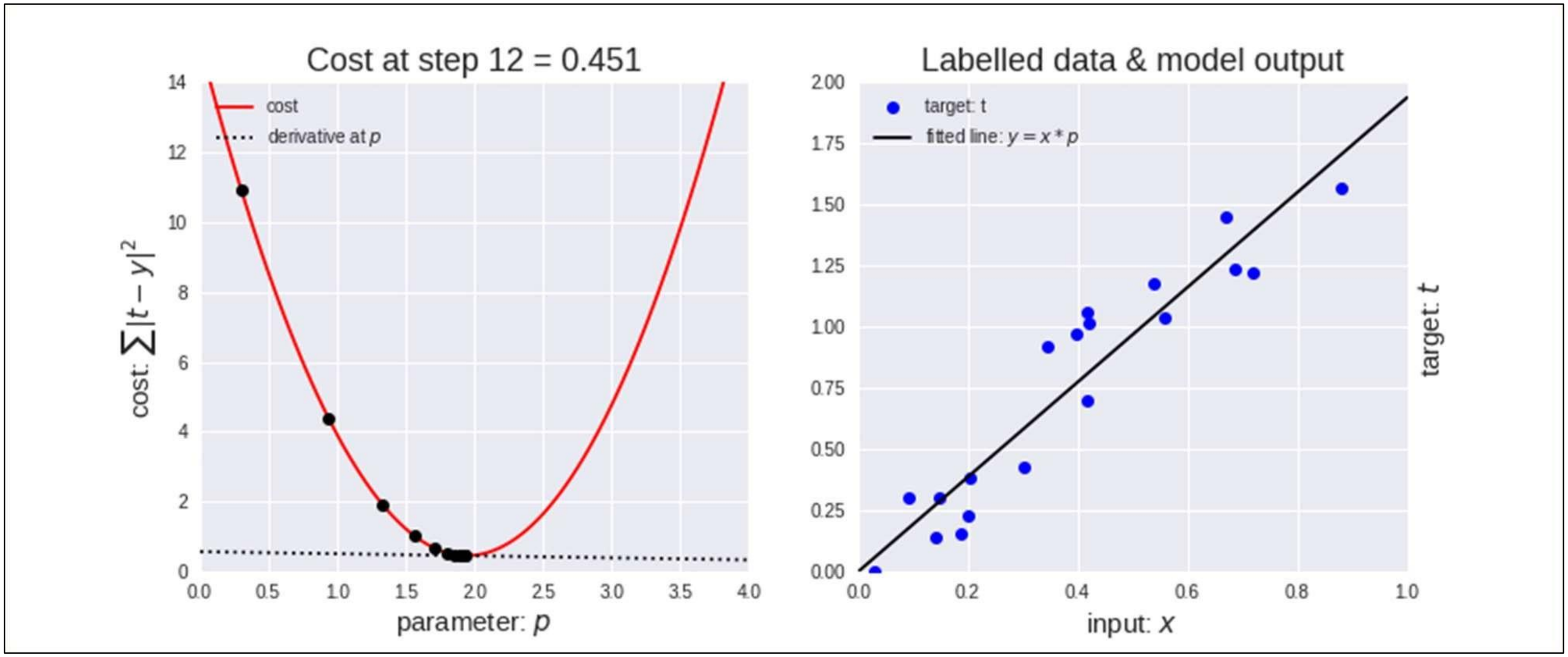
- Step 1: iteration start at x^{k}
- Step 2: search direction

$s^{\mathrm{k}} = negative\ of\ the\ gradient\ of\ f\ x(\ ) = -\nabla f$

- Step 3: new point $\boldsymbol{x^{k+1}} = \boldsymbol{x^{k}} + \boldsymbol{\alpha^{k} s^{k}}$ where α^{k} is the step size



Gradient descent illustration



Q: Find the values of minima found after 3 iteration for function for f(x) given as $f(x) = {x_1}^2$- $2x_1x_2$ + $2x_1$ with constant step size of 0.5 and initial value =(0,0)

- The function is $\mathrm{f}(x_1, x_2) = {x_1}^2 - 2x_1x_2 + 2{x_2}^2 + 2x_1$
- Finding $\nabla f = \begin{bmatrix} 2x_1 - 2x_2 + 2 \\ -2x_1 + 4x_2 \end{bmatrix}$
- The initial values are x^k=(0,0) and step size α=0.5

$$x^1 = x^0 + \alpha s^0 = x^0 - \alpha \nabla f = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.5 \begin{bmatrix} 2(0) - 2(0) + 2 \\ -2(0) + 4(0) \end{bmatrix} = \begin{bmatrix} \mathbf{-1} \\ \mathbf{0} \end{bmatrix}$$

$$x^2 = x^1 + \alpha s^1 = x^1 - \alpha \nabla f = \begin{bmatrix} -1 \\ 0 \end{bmatrix} - 0.5 \begin{bmatrix} 2(-1) - 2(0) + 2 \\ -2(-1) + 4(0) \end{bmatrix} = \begin{bmatrix} \mathbf{-1} \\ \mathbf{-1} \end{bmatrix}$$

$$x^3 = x^2 + \alpha s^2 = x^2 - \alpha \nabla f = \begin{bmatrix} -1 \\ -1 \end{bmatrix} - 0.5 \begin{bmatrix} 2(-1) - 2(-1) + 2 \\ -2(-1) + 4(-1) \end{bmatrix} = \begin{bmatrix} \mathbf{-2} \\ \mathbf{-1} \end{bmatrix}$$

14

More about Hessian

https://web.stanford.edu/group/sisl/k12/optimization/MO-unit4-pdfs/4.10applicationsofhessians.pdf