1] $f(x) = (x_1-4)^2 + (x_2-3)^2 \rightarrow \max$

$g(x)$ $x_1^2 + 2x_2^2 - 1 = 0$

$f(x) + \lambda g(x) \Rightarrow \max$

$(x_1-4)^2 + (x_2-3)^2 + \lambda x_1^2 + \lambda 2x_2^2 - \lambda \Rightarrow \max$

$\frac{\partial f}{\partial x_1} = 2x_1 - 8 + 2\lambda x_1 = 0 \Rightarrow 2x_1 + 2\lambda x_1 = 8 \Rightarrow 2x_1(1+\lambda) = 8$

$x_1 = \frac{8}{2(1+\lambda)} \Rightarrow x_1 = \frac{4}{1+\lambda}$ (1)

$\frac{\partial f}{\partial x_2} = 2x_2 - 6 + 4\lambda x_2 = 0 \Rightarrow 2x_2 + 4\lambda x_2 = 6 \Rightarrow 2x_2(1+2\lambda) = 6$

$2x_2 = \frac{6}{1+2\lambda} \Rightarrow x_2 = \frac{3}{1+2\lambda}$ (2)

(1) & (2) into $g(x)$ to find λ

$\left(\frac{4}{1+\lambda}\right)^2 + 2\left(\frac{3}{1+2\lambda}\right)^2 - 1 = 0 \Rightarrow \frac{16}{(1+\lambda)^2} + \frac{18}{(1+2\lambda)^2} = 1 \Rightarrow$

*I Can't Solve for λ, but next Steps are:

(1) Solve for λ

(2) Plug λ into equations (1) & (2) to find x_1 & x_2

(3) Pick x_1 & x_2 for which $f(x)$ is maximized

2] $x_1 = 0$ $\quad f = 2(x_1+2)^2 + 3(x_2-2)^2 \quad \alpha = 0.1$

$x_2 = 0$

$$\frac{\partial f}{\partial x_1} = 4x_1 + 8$$

$$\frac{\partial f}{\partial x_2} = 6x_2 - 12$$

$$x' = x_i - \alpha * \frac{\partial f}{\partial x_i}$$

$$x_1' = 0 - (0.1 * (4 * 0) + 8) = 0 - 0.8 = -0.8$$

$$x_2' = 0 - (0.1 * (6 * 0) - 12) = 0 + 1.2 = 1.2$$

$$\boxed{\begin{aligned} x_1' &= -0.8 \\ x_2' &= 1.2 \end{aligned}}$$

} First step of gradient descent

Jordan Cook 7739

3.3 What is the Main objective of the traditional NN?

* The main objective of the traditional Neural Network is to speed up computing time & speed up data processing

ie: get all computations OF F as fast as possible

Jordan Caylor | 7739

3.6 Why L-NL-L & not NL-L-NL

① Main objective of traditional neural networks is fast computing.

② NL layers are slower computations than Linear layers.

③ Only need 1 Non linear (NL) layer to approximate any function

with ① ② & ③ we show that L-NL-L is best because it works well for what we want & it is faster than NL-L-NL. Also, fewer layers means faster computing therefore 1 NL layer is fastest.

Jordan Caylor | 7739

3.9

When training a Neural Network we want to minimize the Sum of the Squares of the errors between the given pattern ($\tilde{Y}$) & the calculated pattern (Y)

$$f(x_1^{(p)}, \ldots, x_n^{(p)}) = \tilde{y}^{(p)}$$

$$y^{(p)} = \sum_{k=1}^{ic} W_{ic} * S_k \left(\sum_{i=1}^{n} W_{ki} * x_i - W_{k0} \right) - W_0$$

$$\text{Error} \overset{\text{def}}{\equiv} E$$

$$E = \sum_{p=1}^{P} \left(\tilde{y}_p - y_p \right)^2$$

* Each iteration of the neural network should reduce E.

Jordan Carlson 7739

3.12 We want to minimize J

$$J = \left(\sum_{k=1}^{K} W_k \cdot S\left(\sum_{i=1}^{n} W_{ki} x_i - W_{k0}\right) - W_0\right) - Y)^2 \rightarrow \min \qquad E = Y - \hat{Y}$$

$$W_0 \rightarrow W_0 + \Delta W_0 \; ; \; \Delta W_0 = -\alpha \frac{\delta J}{\delta W_0}$$

$$\frac{\delta J}{\delta W_0} = 2E \frac{\partial E}{\delta W_0} = -2E \quad \therefore \boxed{\Delta W_0 = 2\alpha E}$$

$$\frac{\delta J}{\delta W_{ik}} = 2E \frac{\delta E}{\delta W_{ik}} = 2EY_k \quad \therefore \Delta W_k = -2E X Y_k = -\Delta W_0 Y_k$$

$$\boxed{\begin{aligned} \Delta W_0 &= 2\alpha E \\ \Delta W_k &= -\Delta W_0 Y_k \end{aligned}}$$

Jordan Cantor | 7739

3.1 Smoothness $\equiv x_1 \approx x_2, \ldots, x_1 \simeq x_{n+1}$; $x_{n-1} - x_n \approx 0$.

we want

$$(x_0 - x_1)^2 + \ldots + (x_n - x_{n-1})^2 \to \min$$

$$f \overset{\text{def}}{\equiv} \ldots + (x_i - x_{i-1})^2 + (x_{i+1} - x_i)^2 + \ldots$$

$$f' = 2(x_i - x_{i-1}) + 2(x_{i+1} - x_i)$$

$$x_i - x_{i-1} = x_{i+1} - x_i \} \text{ Straight line}$$

Jordan Caylor 7739

2.2 Machine learning is a way to speed up data processing

We know inputs: $X^{(p)} = (x_1^{(p)}, \ldots, x_n^{(p)})$ & $y^{(p)}$

$1 \le p \le P$

We want: a function

$$f(x_1^{(p)}, \ldots, x_n^{(p)})$$

Such that

$$f(x_1^{(p)}, \ldots, x_n^{(p)}) \approx y^{(p)}$$

Jordan Caylor 7734

3.4 ① 1 nonlinear layer is sufficient to approximate any nonlinear function

② We want the process to be as fast as possible & fewer neurons means ~~fewer~~ less time

with ① & ② it is clear that 1 layer is sufficient

Jordan Caylor | 7739

3.7 All functions can be expressed as a combination of Sinusoidal functions

ie: Newtons prism showed all light can be decomposed into sinusoids corresponding to other light colors (wavelengths).

math:

$$Y = \sum_{k=1}^{K} W_k * S_k \left(\sum_{i=1}^{n} W_{ki} * X_i - W_{k0} \right) - W_0$$

$$F(x_1) \approx \sum_k \left(A_k \sin(w_k * x_1) \mid B_k * \cos(w_k * x_1) \right)$$

$$F(x, y) = \sum_k A_k(x) \sin(w_k y) + B_k(y) \cos(w_k y)$$

$$A_n(x) = \sum A_{ki} * \sin(w_i x) + B_{xi} \cos(w_i x) \sin(w_k y) + \ldots$$

ie: $\cos(a)\cos(b) = \frac{1}{2} * \cos(a+b) + \cos(a-b)$

$$\therefore Y = \sum_{i=1}^{k} W_k * S_k \left(\sum_{i=1}^{n} W_{ki} X_i - W_{k0} \right) - W_0$$

Jordan Taylor

7739

3.5 ① With one or two linear layers we only get linear functions

② With one L & one NL only get functions of one variable

L→NL:

$$S_0\left(\sum_{i=1}^{n} W_i * x_i - W_0\right) \Rightarrow F(x_1, \ldots, x_n) = \text{Const}$$

Jordan Caylor 7739

4

① rescaling: $x \to \lambda x$

② Change in start: $x \to x + x_0$

} $x \to ax + b$] linear functions

③ Need nonlinear transformation

Want: class of natural transformations

- if f & g are natural, $f \circ g$ is natural & reasonable
- Linear transformations are reasonable

Theorem: Every reasonable transformation is fractional linear

$$f(x) = \frac{ax+b}{cx+d}$$

Want: Activation function that won't change if we change starting point

$y = S(x)$

$S(x+x_0) = f(S(x))$ for every reasonable transformation $f(z)$

$$S(x+x_0) = \frac{a(x_0) * S(x) + b(x_0)}{c(x_0) + S(x) + d(x_0)}$$ } functional equation

$$S'(x+x_0) = \frac{a'(x_0) \cdot S(x) + b'(x_0))^* (c(x_0) * S(x) + d(x_0)) - (a(x_0)^* S(x) + b(x_0)) * (C'(x_0)^* S(x) + d'(x_0))}{(C(x_0) * S(x) + d(x_0))^2}$$

$$S(x+y_0) = \frac{a(y_0) * S(x) + b(x_0)}{c(x) * S(x) + 1}$$

$$\frac{dS}{dx} = S'(x) = \overset{A}{a'(0)} * S(x) + \overset{B}{b'(0)} - S(x)^* \overset{C}{c'(0)} * S'(x)$$

$$= A * S(x) + B - C * S^2(x)$$

$$B + As - Cs^2 \Rightarrow \underbrace{\int \frac{ds}{B + As + Cs^2} = \int dx}_{\text{Sigmoid}}$$

Jordan Taylor 7739