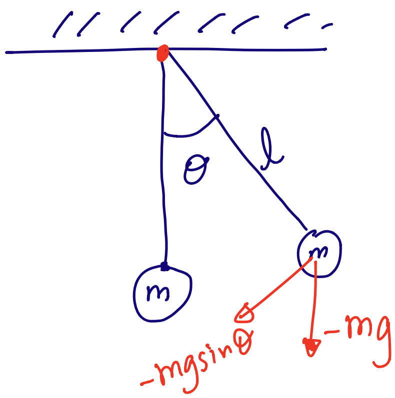
EECS 16 B    Module 3, Lecture 6.    Apr 20, 2021.

Linearization Payoffs.

- Pendulum linearization / Stability.
- Classification – Shortcomings of linear classifiers.
- Logistic Regression – Standard Machine Learning classification technique
  - ↳ "Linearization" is a key technique.
    - ↳ Use "quadratic approximation"

---




$$\underbrace{\overbrace{ml\frac{d^2\theta}{dt^2}}^{m \times acc}}_{accel} = \overbrace{-kl\frac{d\theta}{dt}}^{frictional\ force} \underbrace{- mg\sin\theta}_{gravity.}$$

↑ angular velocity.

$$\cancel{ml}\cdot\frac{dx_2}{dt} = \frac{-k\cancel{l}}{m}\cdot x_2 - \frac{\cancel{m}g\sin x_1}{l}$$

$x_1 = \theta$

$x_2 = \frac{d\theta}{dt}$

$\frac{dx_1}{dt} = x_2.$

$$\frac{d\vec{x}}{dt} = \begin{bmatrix} dx_1/dt \\ dx_2/dt \end{bmatrix} = \underbrace{\begin{bmatrix} x_2 \\ -\frac{g}{l}\sin x_1 - \frac{k}{m}x_2 \end{bmatrix}}_{f(x_1,\, x_2).}$$

Linearize — Operating point

$(x_1, x_2) = (0, 0)$

$(x_1, x_2) = (\pi, 0)$

$\to$ First operating point: $(0, 0)$

$$f(x_1, x_2) \approx f(x_1^*, x_2^*) + \left.\frac{\partial f}{\partial x_1}\right|_{x_1 = x_1^*,\, x_2 = x_2^*} (x_1 - x_1^*)$$

$$+ \left.\frac{\partial f}{\partial x_2}\right|_{x_2 = x_2^*,\, x_1 = x_1^*} (x_2 - x_2^*)$$

$$= \underbrace{f(0, 0)}_{=0} + \left.\left(-\frac{g}{l}\cos x_1 - 0\right)\right|_{\underbrace{x_1 = 0,\, x_2 = 0}} (x - x_1^*)$$

Instead: $x_1^* = \pi, x_2 = 0$

$$+ \left.\left(0 - \frac{k}{m}\right)\right|_{\underbrace{x_1 = 0,\, x_2 = 0}} (x_2 - x_2^*)$$

$$= 0 - \frac{g}{l}\cos 0\,(x_1 - 0) - \frac{k}{m}(x_2 - 0)$$

$$= -\frac{g}{l} \cdot x_1 - \frac{k}{m} x_2 .$$

linearized eq about operating point $(0, 0)$

$$\frac{d\vec{x}}{dt} = \begin{bmatrix} x_2 \\ -\frac{g}{l} x_1 - \frac{k}{m} x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -\frac{g}{l} & -\frac{k}{m} \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\det(\lambda I - A) = \det\left(\begin{bmatrix} \lambda & -1 \\ g/l & \lambda + \frac{k}{m} \end{bmatrix}\right)$$

$$= \lambda\left(\lambda + \frac{k}{m}\right) + g/l$$

$$= \lambda^2 + \frac{k}{m}\lambda + g/l.$$

$$\boxed{k, m, g, l > 0}$$

Roots:

$$\lambda = \frac{\underbrace{-\left(\frac{k}{m}\right)}_{\text{neg}} \pm \sqrt{\frac{k^2}{m^2} - 4 g/l}}{2}$$

always has negative real parts!

---

Consider other operating point $(\pi, 0)$

Linearized eq$^n$:

$$f(x_1, x_2) \approx f(\pi, 0) - \frac{g}{l}\cos\pi\,(x_1 - \pi) - \frac{k}{m}\cdot x_2.$$

$$\frac{d\vec{x}}{dt} = \begin{bmatrix} 0 & 1 \\ g/l & -\frac{k}{m} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \text{constant}$$

$$\det\begin{bmatrix} \lambda & -1 \\ -g/l & \lambda + \frac{k}{m} \end{bmatrix} = \lambda(\lambda + k/m) - g/l.$$

$$= \lambda^2 + \frac{k}{m}\lambda - g/l$$

Roots:
$$\frac{-\left(\frac{k}{m}\right) \pm \sqrt{\frac{k^2}{m^2} + 4(g/l)}}{2}.$$

$$\frac{k^2}{m^2} + \frac{4g}{l} > 0 \Rightarrow \text{ roots are always real}$$

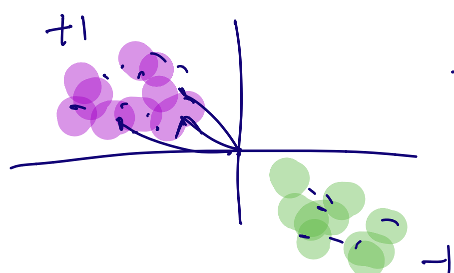
$$\sqrt{\frac{k^2}{m^2} + \frac{4g}{l}} > \sqrt{\frac{k^2}{m^2}}$$

$$-\frac{k}{m} + \sqrt{\frac{k^2}{m^2} + \frac{4g}{l}} > 0$$

$\Rightarrow$ Always one e-val $> 0$.

---

Classification




$$l_i = \underbrace{\{+1, -1\}}_{\text{labels}}.$$

$(\vec{x}_1, l_1)$ $(\vec{x}_2, l_2)$ ... Data points.

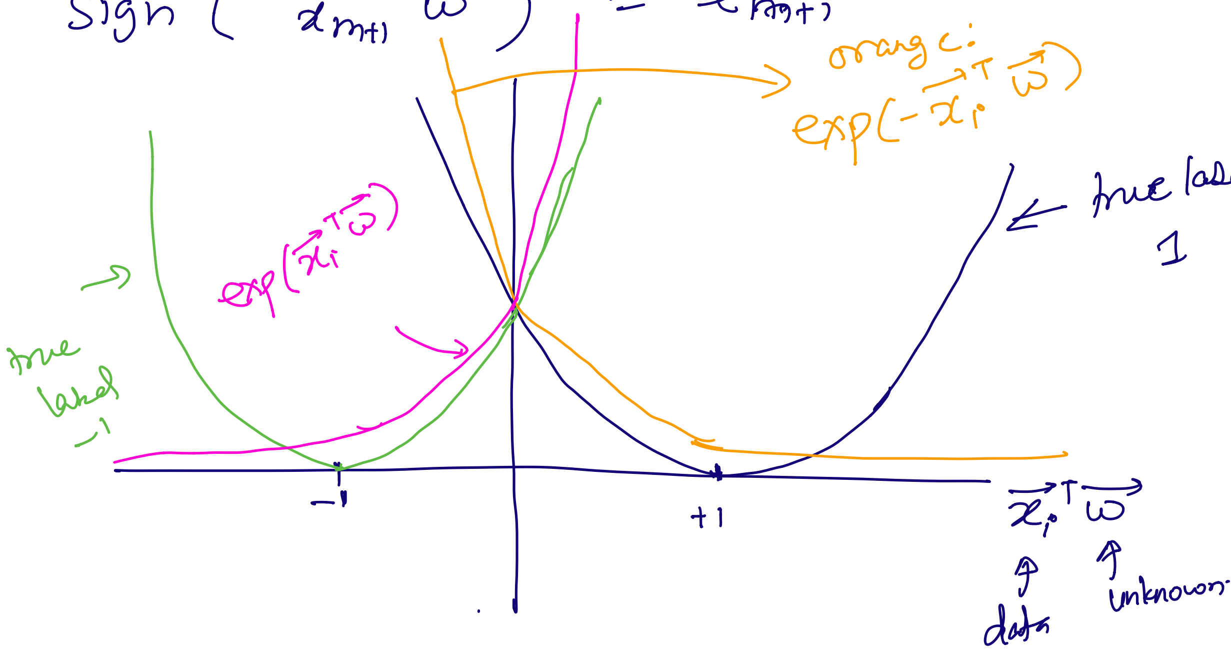
Augmented: $\vec{x}_1 \rightarrow \begin{bmatrix} 1 \\ \vec{x}_1 \end{bmatrix}$. $\triangleq$ $\begin{bmatrix} \vec{x}_1^T & 1 \\ \vec{x}_2^T & 1 \\ \vdots & \\ \vec{x}_m^T & 1 \end{bmatrix} \begin{bmatrix} \vec{w} \\ b \end{bmatrix} - \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_m \end{bmatrix}$

Least squares cost.

$$c(\vec{x}_i^T \vec{w}, l_i) = \|\vec{x}_i^T \vec{w} - l_i\|^2$$

"Right" $\vec{w}$ gives me a classifier

$$\text{sign}\left(\vec{x}_{m+1}^T \vec{w}\right) = l_{m+1}$$




When true label is +1, I want

$$\text{sign}\left(\vec{x}_i^T \vec{w}\right) > 0$$

When true label is $-1$, want $\text{sig}\left(\vec{x}_i^T \vec{w}\right) < 0$

Cost function of choice:

$$\exp\left(-l_i \vec{x}_i^T \vec{w}\right)$$

---

Instead of the quadratic cost:

$$C\left(\vec{x}_i^T \vec{w}, l_i\right) = \exp\left(-l_i \vec{x}_i^T \vec{w}\right)$$

cost function.

Problem: Find $\vec{w}$ so that

$$\underset{\vec{w}}{\operatorname{argmin}} \underbrace{\sum_{i=1}^{m} \exp(-l_i \vec{x}_i^T \vec{w})}$$

$l_i$, $\vec{x}_i$: training data

$\vec{w}$: unknown wants to find

Least-squares:

$$\underset{\vec{w}}{\operatorname{argmin}} \sum_{i=1}^{m} \|\vec{x}_i^T \vec{w} - l_i\|^2$$

BAD

Better

$$f(w) \approx f(w^*) + f'(w^*)(w - w^*) + \frac{1}{2} f''(w)(w - w^*)^2$$

$f(\vec{w})$

If $\vec{w}$ is a vector:

1st Derivative of $f$ w.r.t. $\vec{w}$ is a row vector.

2nd derivative : derivative of a vector: Matrix.

$$f(\vec{w}) \approx f(\vec{w}^*) + \overbrace{\begin{bmatrix} \frac{\partial f}{\partial w_1} & \frac{\partial f}{\partial w_2} & \cdots & \frac{\partial f}{\partial w_n} \end{bmatrix}}^{\text{1st derivative}}\Bigg|_{\vec{w}=\vec{w}^*} (\vec{w} - \vec{w}^*)$$

$$+ \frac{1}{2} (\vec{w} - \vec{w}^*)^T \begin{bmatrix} \frac{\partial^2 f}{\partial w_1^2} & \frac{\partial^2 f}{\partial w_1 \partial w_2} & \cdots & \frac{\partial^2 f}{\partial w_n \partial w_1} \\ \vdots & & \vdots & \\ \frac{\partial^2 f}{\partial w_1 \partial w_n} & & \cdots & \frac{\partial^2 f}{\partial w_n^2} \end{bmatrix}\Bigg|_{\vec{w}=\vec{w}^*} (w - \vec{w}^*)$$

$$\frac{\partial^2 f}{\partial w_1 \partial w_2} = \frac{\partial^2 f}{\partial w_2 w_1}$$ f sufficiently continuous.

Symmetric Matrix.

"Hessian"

---

Office hours

$$\begin{bmatrix} \vec{x}_1^T & 1 \\ \vdots & \\ \vec{x}_m^T & 1 \end{bmatrix} \begin{bmatrix} \vec{w} \\ b \end{bmatrix} = \begin{bmatrix} l_1 \\ \vdots \\ l_m \end{bmatrix}.$$ least squ

$$\overbrace{\begin{bmatrix} - \vec{x}_1^T - \\ \vdots \\ - \vec{x}_m^T - \end{bmatrix}}^{A} \vec{w} = \underbrace{\begin{bmatrix} l_1 \\ \vdots \\ l_m \end{bmatrix}}_{\vec{l}}$$

$$\underset{\vec{w}}{\operatorname{argmin}} \left[ (\vec{x}_1^T \vec{w} - l_1)^2 + (\vec{x}_2^T \vec{w} - l_2)^2 \cdots + (\vec{x}_m^T \vec{w} - \vec{l}_m)^2 \right]$$

$$\updownarrow$$

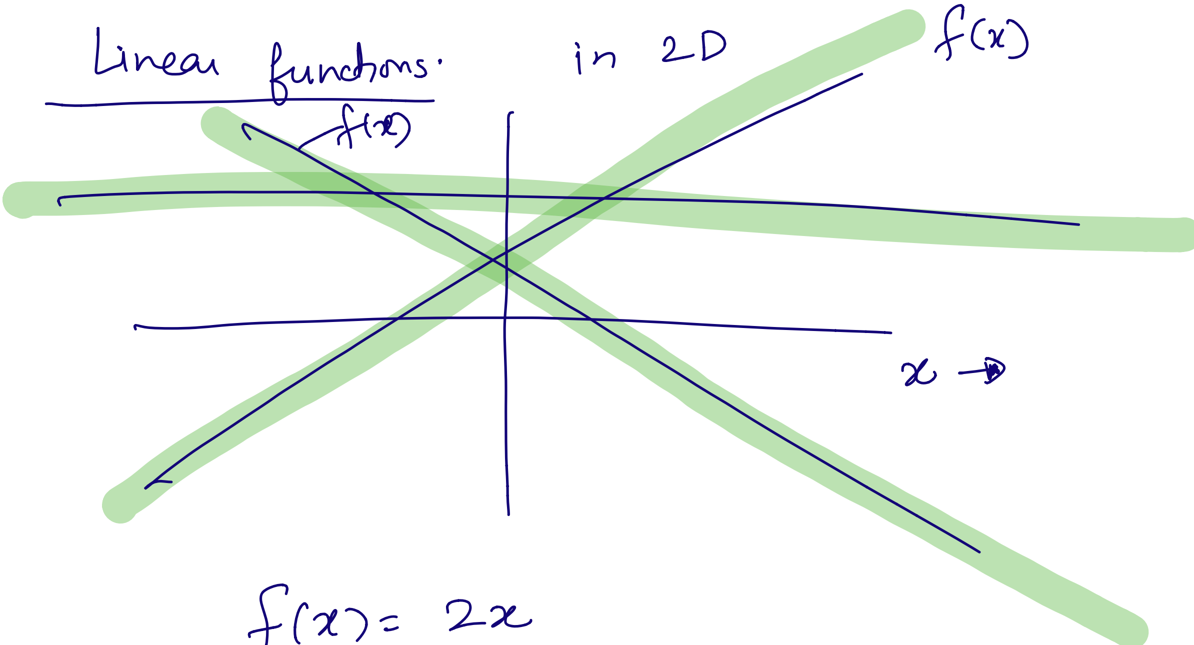
$$\underset{\vec{w}}{\operatorname{argmin}}: \quad \|A\vec{w} - \vec{l}\|^2$$ "cost function"

$$\underset{\vec{w}}{\operatorname{argmin}} \sum_{i=1}^{m} \underbrace{(\vec{x_i}^T \vec{w} - l_i)^2}_{\text{cost}(\vec{x_i}^T \vec{w},\ l_i)}$$

$$= \underset{\vec{w}}{\operatorname{argmin}} \sum_{i=1}^{m} c(\vec{x_i}^T \vec{w}, l_i)$$

---

Linear functions. in 2D




$f(x) = 2x$

$\min_x f(x) = -\infty$

$\underset{x}{\operatorname{argmin}} f(x) = -\infty$

$f(x) = -2x$

$\min_x f(x) = -\infty$

$\underset{x}{\operatorname{argmin}} f(x) = +\infty$

$f(x) = 5$