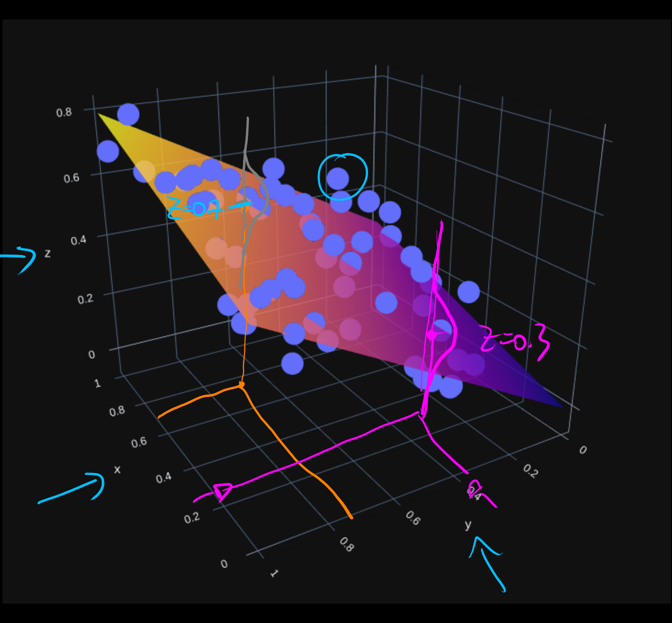
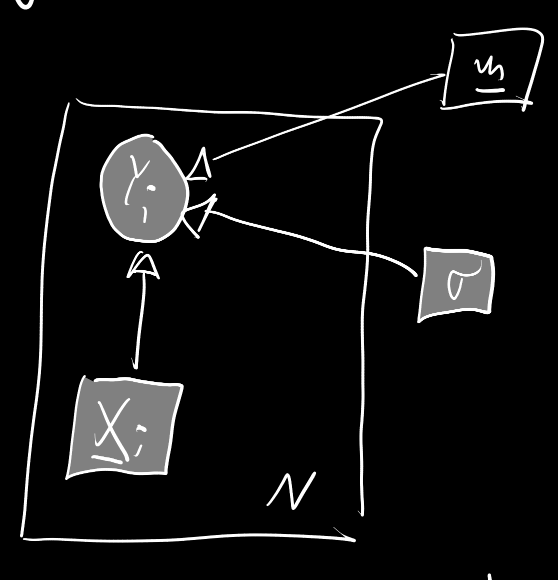
Linear Regression in High Dimensions



$$\underline{w}^* = \underset{\underline{w} \in \mathbb{R}^D}{\arg\min} \left(\frac{1}{2} \| \underline{Y} - \underline{\underline{X}}\,\underline{w} \|_2^2 \right)$$

Agenda:

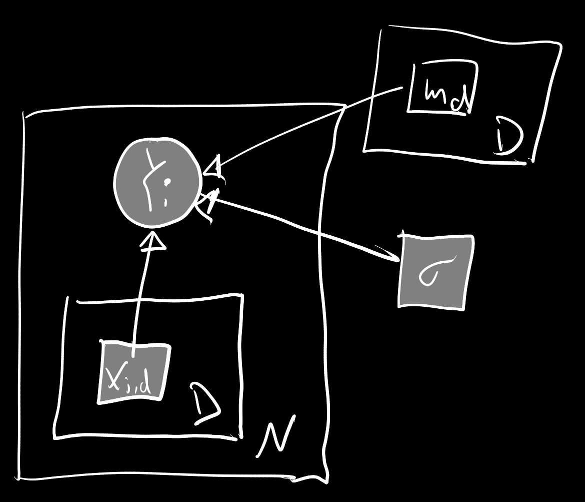
1. Directed Graphical Model
2. Least Squares via univariate Normal
3. Least Squares via multivariate Normal
4. Demo in Python



$Y_i \sim$ Normal

$$Y_i \sim \mathcal{N}(Y_i, \mu = \underline{X}_i^T \underline{w}, \sigma)$$

equally:



$$p(Y_i | \underline{X}_i, \underline{w}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (Y_i - \underbrace{\underline{X}_i^T \underline{w}}_{= \mu \in \mathbb{R}})^2 \right)$$

Likelihood

$$\mathcal{L}(D) \overset{\text{i.i.d.}}{=} \prod_{i=0}^{N-1} p(Y_i = y^{[i]} | \underline{X}_i = \underline{x}^{[i]}, \underline{w}, \sigma)$$

$$= \prod_{i=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y^{[i]} - \underline{x}^{[i]T} \underline{w})^2 \right)$$

$$\propto \prod_{i=0}^{N-1} \exp\left(-\frac{1}{2\sigma^2} (y^{[i]} - \underline{x}^{[i]T} \underline{w})^2 \right)$$

$$= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (y^{[i]} - \underline{x}^{[i]T} \underline{w})^2 \right)$$

Log-Likelihood

$$\ell(D) = \log(\mathcal{L}(D))$$

$$\overset{+}{=} -\frac{1}{2\sigma^2} \sum_{i=0}^{N-1} (y^{[i]} - \underline{x}^{[i]T} \underline{w})^2$$

$$\overset{+}{\propto} -\frac{1}{2} \sum_{i=0}^{N-1} (y^{[i]} - \underline{x}^{[i]T} \underline{w})^2$$

MLE

$$\underline{w}^* = \underset{\underline{w} \in \mathbb{R}^D}{\arg\max} \left(-\frac{1}{2} \sum_{i=0}^{N-1} (y^{[i]} - \underline{x}^{[i]T} \underline{w})^2 \right)$$

$$= \underset{\underline{w} \in \mathbb{R}^D}{\arg\min} \left(\underbrace{\frac{1}{2} \sum_{i=0}^{N-1} (y^{[i]} - \underline{x}^{[i]T} \underline{w})^2}_{t} \right)$$

Least Squares

datamatrix

$$\underline{\underline{X}} = \begin{bmatrix} - & x^{[0]} & - \\ - & x^{[1]} & - \\ - & x^{[2]} & - \\ & \vdots & \end{bmatrix} \Bigg\} N \text{ rows}$$

D columns

parameter vector

$$\underline{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \end{bmatrix} \Big\} D$$

target vector

$$\underline{y} = \begin{bmatrix} y^{[0]} \\ y^{[1]} \\ y^{[2]} \\ \vdots \end{bmatrix} \Bigg\} N \text{ rows}$$

$$\underline{y} - \underline{\underline{X}}\,\underline{w}$$

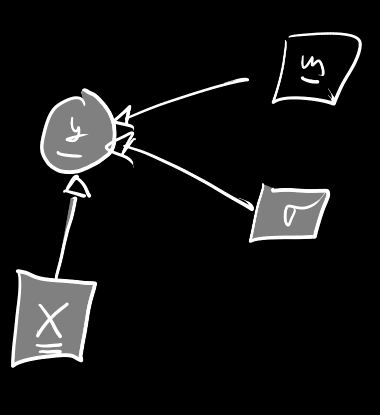
$$t = \frac{1}{2} \sum_{i=0}^{N-1} (\underline{y} - \underline{\underline{X}}\,\underline{w})_i^2$$

$$= \frac{1}{2} \underbrace{(\underline{y} - \underline{\underline{X}}\,\underline{w})^T}_{\Delta \underline{y}} \underbrace{(\underline{y} - \underline{\underline{X}}\,\underline{w})}_{\Delta \underline{y}}$$

$$= \frac{1}{2} \| \underline{y} - \underline{\underline{X}}\,\underline{w} \|_2^2$$

$$\Delta \underline{y} = \begin{bmatrix} \Delta y^{[0]} \\ \Delta y^{[1]} \\ \vdots \end{bmatrix} \Bigg\} N \text{ rows}$$

From the MV Normal



isotropical MV Gaussian

$$\underline{y} \sim \mathcal{N}(\underset{\in \mathbb{R}^N}{\underline{y}}, \underset{\in \mathbb{R}^N}{\underline{\mu}} = \underset{\in \mathbb{R}^{N \times D}}{\underline{\underline{X}}}\,\underset{\in \mathbb{R}^D}{\underline{w}}, \underline{\underline{\Sigma}} = \sigma^2 \underline{\underline{I}})$$

$$p(\underline{y} | \underline{\underline{X}}, \underline{w}, \sigma) =$$

$$= \frac{1}{\sqrt{(2\pi\sigma^2)^N}} \exp\left(-\frac{1}{2\sigma^2} (\underline{y} - \underline{\underline{X}}\,\underline{w})^T (\underline{y} - \underline{\underline{X}}\,\underline{w}) \right)$$

Likelihood

$$\mathcal{L}(D) = p(\underline{y} | \underline{\underline{X}}, \underline{w}, \sigma)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2} (\underline{y} - \underline{\underline{X}}\,\underline{w})^T (\underline{y} - \underline{\underline{X}}\,\underline{w}) \right)$$

Log-Likelihood

$$\ell(D) = \log(\mathcal{L}(D))$$

$$\overset{+}{\propto} -\frac{1}{2} (\underline{y} - \underline{\underline{X}}\,\underline{w})^T (\underline{y} - \underline{\underline{X}}\,\underline{w})$$

$$= -\frac{1}{2} \| \underline{y} - \underline{\underline{X}}\,\underline{w} \|_2^2$$

$$\underline{w}^* = \underset{\underline{w} \in \mathbb{R}^D}{\arg\min} \left(\frac{1}{2} \| \underline{y} - \underline{\underline{X}}\,\underline{w} \|_2^2 \right)$$

standard problem in linear algebra

→ linear least squares problem

For example with NumPy

np.linalg.lstsq($\underline{\underline{A}}$, $\underline{b}$) $\longrightarrow \underline{w}^*$

→ There is a closed-form solution

→ no numerical optimization is needed (no gradient descent)