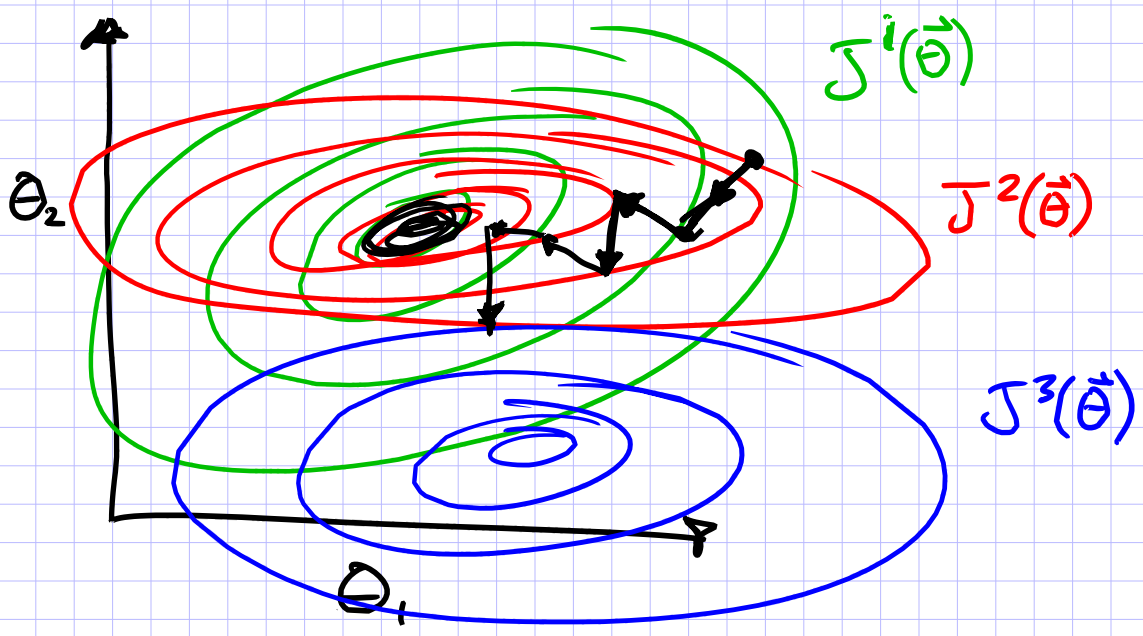
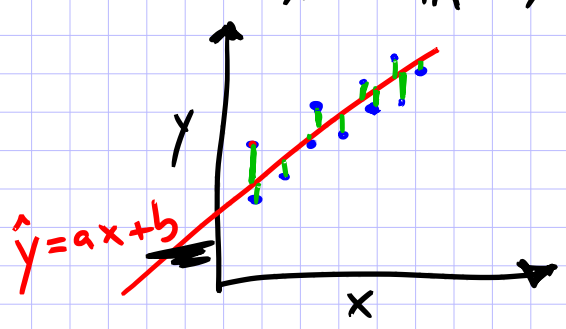
Jelinek Summer School



Linear Reg.

Data $D = \{(\vec{x}^{(i)}, y^{(i)})\}_{i=1}^{N} = \{(\vec{x}^{(1)}, y^{(1)}), (\vec{x}^{(2)}, y^{(2)}), \dots\}$

$\vec{x}^{(i)} \in \mathbb{R}^M, \quad y^{(i)} = \mathbb{R}$



a, b parameters.

Linear Fn.

$$y = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_M x_M + \theta_0$$

$$y = \theta_0 + \sum_{m=1}^{M} \theta_m x_m = \underline{\theta_0} + \vec{\theta}^T \vec{x}$$

$$y = \vec{\theta}^T \vec{x} \quad \text{if } x_0 = 1$$

$$\boxed{\frac{dh(\vec{x})}{d\theta_m} = x_m}$$

Lin. Reg. as Fn Approx.

① Assume D generated $\vec{x}^{(i)} \sim p^*(\vec{x})$

$\vec{y}^{(i)} = h^*(x^{(i)})$ unk.

② Choose Hyp. Space, H

$$H = \{h(\vec{x}) = \vec{\theta}^T \vec{x} : \vec{\theta} \in \mathbb{R}^M\}$$

space of all lin. fns in M-dim

$$\frac{dJ^{(i)}(\theta)}{d\theta_m} = \frac{2}{N}(y^{(i)} - \theta^T x^{(i)}) x_m^{(i)}$$

$O(M)$

③ Choose Objective Fn.

$$J(\vec{\theta}) = \sum_{i=1}^{N} \underbrace{\frac{1}{N}(y^{(i)} - \vec{\theta}^T \vec{x}^{(i)})^2}_{J^{(i)}(\theta)}$$

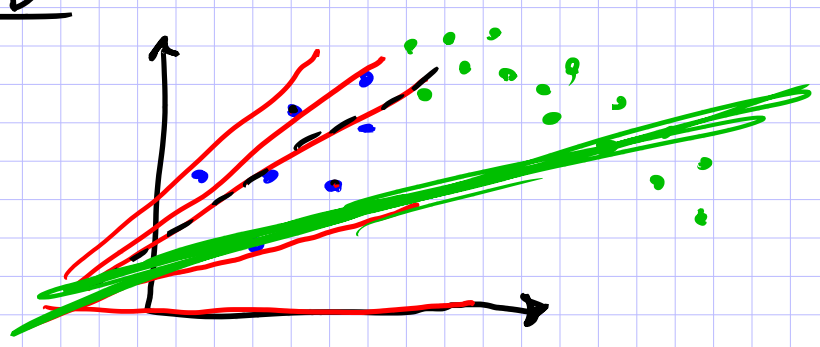
$$\text{MSE} = \sum_{i=1}^{N}$$

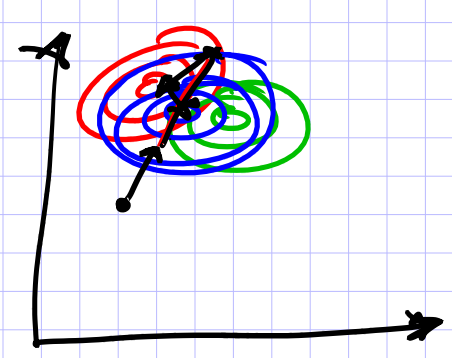
④ Solve unconstrn opt. prob. (SGD, GD, closed-form)

$$\hat{\theta} = \arg\min J(\vec{\theta})$$

⑤ Predict: $\hat{y} = h(\vec{x}) = \vec{\theta}^T \vec{x}$

SGD:





$O(N^{2.783\ldots})$

Closed Form:

$$\vec{\theta} = (X^T X)^{-1}(X^T \vec{y})$$

$N \times M$

$$\vec{y} = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{bmatrix} \quad X = \begin{bmatrix} \text{---} \; \vec{x}^{(1)} \; \text{---} \\ \vdots \\ \text{---} \; x^{(N)} \; \text{---} \end{bmatrix}$$

Log. Reg.

$$p(D) = \left[\prod_{i=1}^{N/2} \underbrace{p(y^{(i)}|\vec{x})}_{1}\right]\left[\prod_{i=N/2+1}^{N} \underbrace{p(y^{(i)}|\vec{x})}_{1}\right]$$

$$= \left[\prod_{i=1}^{N/2} \underbrace{1-p(y=1|\vec{x})}_{1}\right]\left[\prod_{i=N/2+1}^{N} \underbrace{p(y=1|\vec{x})}_{1}\right]$$

$$1 = p(y=0|\vec{x}^{(i)}) + p(y=1|\vec{x}^{(i)})$$

$$J(\theta) = -\log p(D)$$
$$= -\log \prod_{i=1}^{N} p(y^{(i)}|\vec{x}^{(i)})$$
$$= \sum_{i=1}^{N} \underbrace{-\log p(y^{(i)}|x^{(i)})}_{J^{(i)}(\vec{\theta})}$$
$$= \sum_{i=1}^{N}$$

$K=3$

$$\vec{\theta} = \begin{bmatrix} - & \vec{\theta}_1 & - \\ - & \vec{\theta}_2 & - \\ - & \vec{\theta}_3 & - \end{bmatrix}$$

$$\vec{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{bmatrix} \begin{matrix} \vec{\theta}_1 \\ \vec{\theta}_2 \\ \vec{\theta}_3 \end{matrix}$$

Multinomial Logistic Regression

Data: $D = \{(\vec{x}^{(i)}, y^{(i)})\}$ $\quad y^{(i)} \in \{1, \dots, K\}$

Model:
$$p(y|\vec{x};\vec{\theta}) = \frac{\exp(\vec{\theta}_y^T \vec{x})}{Z(\vec{x})}$$
$$= \frac{\exp(\vec{\theta}_y^T \vec{x})}{\sum_{y'=1}^{K} \exp(\vec{\theta}_{y'}^T \vec{x})}$$

Objective:
$$J(\vec{\theta}) = \sum_{i=1}^{N} J^{(i)}(\vec{\theta}) \quad \text{where} \quad J^{i}(\vec{\theta}) = -\log p(y^{(i)}|\vec{x}^{(i)})$$

Gradient:

$$\nabla J^{(i)}(\vec{\theta}) = \begin{bmatrix} \leftarrow \\ \leftarrow \\ \leftarrow \end{bmatrix}$$

$$\frac{dJ^{(i)}(\vec{\theta})}{d\theta_m} = \frac{d}{d\theta_m}\left(-\log p(y^{(i)}|x^{(i)})\right)$$

$$\frac{d}{d\theta_m}\left[-\log p(y|\vec{x};\vec{\theta})\right] = ?$$

$$\frac{d}{d\theta_{zm}}\left[-\log p(y|\vec{x};\vec{\theta})\right] = \frac{d}{d\theta_{zm}}\left[-\log\left(\frac{\exp(\theta_y^T\vec{x})}{Z(\vec{x};\vec{\theta})}\right)\right]$$

$$= \frac{d}{d\theta_{zm}}\left(-\left[\theta_y^T x - \log Z(\vec{x};\theta)\right]\right)$$

$$= \frac{d}{d\theta_{zm}}\left(-\theta_y^T x + \log \underbrace{\sum_{y'} \exp(\theta_{y'}^T x)}\right)$$

$$= \frac{d}{d\theta_{zm}}\left(-\sum_{n=1}^{M}\theta_{yn}x_n + \log Z(\,)\right)$$

$$= -\mathbb{I}(y=z)\,x_m + \frac{d}{d\theta_{zm}}\left(\log Z(\,)\right)$$

$$\frac{d}{d\theta_{zm}}\log Z(\vec{x};\vec{\theta}) = \frac{1}{Z(\vec{x};\theta)}\cdot\frac{d}{d\theta_{zm}}\sum_{y'}\exp(\theta_{y'}^T\vec{x})$$

$$= \frac{1}{Z(\vec{x};\theta)}\cdot\frac{d}{d\theta_{zm}}\exp(\theta_z^T\vec{x})$$

$$= \frac{\exp(\theta_z^T x)}{Z(\vec{x};\theta)}\cdot\frac{d}{d\theta_{zm}}(\theta_z^T\vec{x})$$

$$= \frac{\exp(\theta_z^T x)}{Z(\vec{x};\theta)}\cdot x_m$$

$$= p(z|\vec{x};\vec{\theta})\,x_m$$

$$= -\mathbb{I}(y=z)\,x_m + p(z|\vec{x};\vec{\theta})\,x_m$$

L2 Regularization

$$J(\vec{\theta}) = \left[\frac{1}{N}\sum_{i=1}^{N} J^{(i)}(\vec{\theta})\right] + \lambda\sum_{m=1}^{M}\theta_m^2$$

$$= \left[\frac{1}{N}\sum_{i=1}^{N} -\log p(y^{(i)}|x^{(i)};\vec{\theta})\right] + \lambda\|\vec{\theta}\|_2^2$$