Recap

$$\mathcal{L}(D, \vec{w}, w_0) = -\sum_{D} \left[(\vec{x}\cdot\vec{w} + w_0)\, y \right] + \lambda (\vec{w}\cdot\vec{w})$$

y: target output; $\vec{w}$: weight; w_0: bias

↳ minimize (cross function)

λ: hyperparameter ↳ Regularizer

Gradient Descent

$$\Delta_{\vec{w}} \mathcal{L}(D, y, \vec{w}, w_0) = -\sum_{D} y\,\vec{x} + \lambda \frac{\vec{w}}{\|\vec{w}\|}$$

$$\vec{w}^{(k+1)} = \vec{w}^{(k)} - \eta \nabla_{\vec{w}} \mathcal{L}(D, y, \vec{w}, w_0)$$

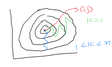
$$= \vec{w}^{(k)} - \eta \left(-\sum_{D} y\,\vec{x} + \lambda \frac{\vec{w}}{\|\vec{w}\|} \right)$$

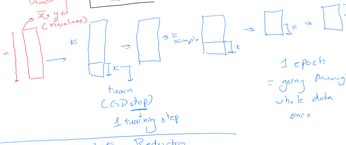
$$w_0^{(k+1)} = w_0^{(k)} - \eta \left(-\sum_{D} y \right)$$

$D^{train} \to n \to$ Gradient Descent

↳ $k = 1$: stochastic gradient descent

↳ $k < n$: Batch gradient descent





1 epoch = going through whole data once

Dimension already Reduction

$$D = \{(\vec{x}_i, y_i) : \vec{x}_i \in \mathbb{R}^{d}\}_{i=1}^{n}$$

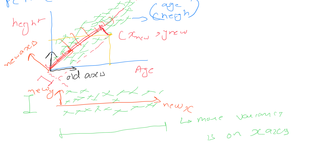
PCA — Principal Component Analysis

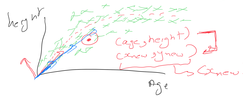
$$D = \{(\vec{x}_i, y_i) : \vec{x}_i \in \mathbb{R}^{d'}\}_{i=1}^{n}$$

$d' < d$

PCA (Unsupervised)

new → information = variance





PCA Implementation

I] standardisation

$$\frac{\vec{x}_i - \mu}{\sigma}$$

$(x_i - \mu)$

II] → $\vec{u}$ → $x' = \dfrac{\vec{x}_i \cdot \vec{u}}{\|u\|}$

↳ objective → high variance

$$Var = \frac{1}{n}(x' - \mu)^2$$

$$= \frac{1}{n}\left(\frac{\vec{x}_i \cdot \vec{u}}{\|u\|} - \cancel{\mu}^{\,0}\right)^2$$

$$= \frac{1}{n}\left(\frac{\vec{x}_i \cdot \vec{u}}{\|u\|}\right)^2$$

st u is a unit vector (condition)

$$Var(x_i) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\vec{x}_i^T \vec{u}}{\|u\|}\right)^2$$

st $\|\vec{u}\|^2 = 1$

$$Var(x_i) = \underset{\vec{u}}{\operatorname{argmax}} \; \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\vec{x}_i^T \vec{u}}{\|u\|}\right)^2 + \lambda(\|u\|^2 - 1)$$

$$\underset{\vec{u}}{\operatorname{argmax}} = \frac{1}{n}\sum_{i=1}^{n}(\vec{x}_i^T \vec{u})^2 + \lambda(\|\vec{u}\|^2 - 1)$$

$$(n \times d)\; X = \begin{bmatrix} \leftarrow x_1^T \rightarrow \\ \leftarrow x_2^T \rightarrow \\ \leftarrow x_3^T \rightarrow \\ \leftarrow x_n^T \rightarrow \end{bmatrix}_{n \times d} \qquad \vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_d \end{bmatrix}$$

→ data vector

$$\boxed{X \cdot \vec{u} = \vec{v}}$$

$$\begin{bmatrix} x_1^T \vec{u} \\ x_2^T \vec{u} \\ x_3^T \vec{u} \\ \vdots \\ x_n^T \vec{u} \end{bmatrix} = \vec{v}$$

$$\vec{u}^* = \underset{\vec{u}}{\operatorname{argmax}} \; \frac{1}{n}(X \cdot \vec{u})^2 + \lambda(\|u\|^2 - 1)$$

$$\vec{v}^T \to \vec{v}^T \vec{v}$$

$$= \underset{\vec{u}}{\operatorname{argmax}} \; \frac{1}{n}(\vec{v}^T \vec{v}) + \lambda(\|u\|^2 - 1)$$

$$\vec{v}^T \vec{v} = (X\vec{u})^T \cdot X\vec{u}$$

$$\vec{u}^* = \underset{\vec{u}}{\operatorname{argmax}} \; \frac{1}{n}\, \vec{u}^T \underbrace{X^T X}_{S} \vec{u} + \lambda(\|\vec{u}\|^2 - 1)$$

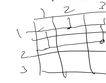
$$f(\vec{u}) = \vec{u}^T S \vec{u}$$

$$\nabla_{\vec{u}} = (S + S^T)\vec{u}$$

$$\boxed{X^T X} \qquad X \;\; n \times d$$

$$d \times n - n \times d$$

$$\boxed{d \times d}$$



$$\vec{u}^* = \underset{\vec{u}}{\operatorname{argmax}} \; \frac{1}{n}\, \vec{u}^T \underbrace{X^T X}_{S} \vec{u} + \lambda(\|\vec{u}\|^2 - 1)$$

$$\vec{u}^T \to \vec{u}^T$$

$$= \underset{\vec{u}}{\operatorname{argmax}} \; \frac{1}{n}\, \vec{u}^T S \vec{u} + \lambda(\|\vec{u}\|^2 - 1)$$

$$\Delta_{\vec{u}} = \frac{1}{n}(S + S^T)\vec{u} + 2\lambda\vec{u}$$

$$= \frac{1}{n}(S + S^T)\vec{u} + 2\lambda\vec{u}$$

$$(X^T X) = S$$

$$= \frac{1}{n}\left((X^T X)^T + X^T X\right)\vec{u} + 2\lambda\vec{u}$$

$$X^T X = (X^T X)^T$$

$$= \frac{1}{n}\cancel{2}\, X^T X \vec{u} + \lambda(\cancel{2}\vec{u}) = 0$$

$$\frac{1}{n} X^T X \vec{u} = -\lambda\vec{u}$$

$$X^T X \vec{u} = (-n\lambda)\,\vec{u}$$

$$\boxed{X^T X}\,\vec{u} = \lambda'\vec{u}$$

→ λ' → eigen value

→ eigen vector $\vec{u}$

$$S\vec{u} = \lambda'\vec{u}$$

↳ eigen vector

eigen vector

$\boxed{S}$ → $(n \times n)$ matrix

$$S\vec{x} = \lambda\vec{x}$$

→ eigen vector