MLF

WEEKS 1-4

1. | Category | Method | Loss Function |
|---|---|---|
| Supervised Learning | Regression | $\frac{1}{n}\sum_{i=1}^{n}(f(x^i)-y^i)^2$ |
| | Classification | $\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(f(x^i)\neq y^i)$ |
| Unsupervised Learning | Dimensionality Reduction | $\frac{1}{n}\sum_{i=1}^{n}\lVert g(f(x^i))-x^i\rVert^2$ |
| | Density Estimation | $\frac{1}{n}\sum_{i=1}^{n}-\log(P(x))$ |

2. Linear approximation of $f(x)$ around x^*
$$L_{x^*}[f](x) = f(x^*) + f'(x^*)(x-x^*)$$

3. Quadratic approximation of $f(x)$ around x^*
$$L_{x^*}[f](x) = f(x^*) + f'(x^*)(x-x^*) + \frac{1}{2}f''(x^*)(x-x^*)^2$$

4. Linear approximation of multivariable scalar function.
$$L_{x^*}[f](x) = f(x^*) + \nabla f(x^*)(x-x^*)$$

5. Projection
   - → b onto a = $\left(\frac{aa^T}{a^Ta}\right)b$
   - → projection matrix → $P = \frac{aa^T}{a^Ta}$
   - → projection onto a subspace → $A^TA\hat{x} = A^Tb$ → $\hat{x} = (A^TA)^{-1}A^Tb$
   - → projection matrix onto a subspace → $P = A(A^TA)^{-1}A^T$

6. Eigenvalue equation → $A\underbrace{x}_{\text{eigenvector}} = \underbrace{\lambda}_{\text{eigenvalue}}x$

7. Characteristic polynomial of matrix A → $\det(A-\lambda I)=0$
   - → solutions of this polynomial gives ==eigenvalues==
   - → $\sum_{i=1}^{n}\lambda_i = \text{trace}(A)$; $\prod_{i=1}^{n}\lambda_i = \det(A)$

8. Eigenvectors calculated using $(A-\lambda I)x=0$, solving for x using λ eigenvalues.

9. Rank of matrix = # of eigenvalues = # of linearly independent eigenvectors.

10. $S^{-1}AS=\Lambda$ → symmetric matrix of eigenvalues in diagonal

11. $S = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix}$, where $\{x_1, x_2, \ldots, x_n\}$ are ==orthogonal eigenvectors== of matrix A.
    - → S is not unique because eigenvectors can be different.
    - → $A^2x = \lambda^2 x$
    - → $S^{-1}A^2x = \Lambda^2$

12. $A = Q\Lambda Q^T$, where Q is orthogonal matrix.

$Q = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix}$, where $\{x_1, x_2, \ldots, x_n\}$ are ==orthonormal eigenvectors== of matrix A.

WEEKS 5-8

① Complex matrices , $x, y \in \mathbb{C}^n$

- $\rightarrow$ $x \cdot y = \bar{x}^T \cdot y$
- $\rightarrow$ $x \cdot y = \overline{y \cdot x}$
- $\rightarrow$ $cx \cdot y = \bar{c}(x \cdot y)$
- $\rightarrow$ $(AB)^* = B^* A^*$

② Hermitian matrices

- $\rightarrow$ $A^* = A$
- $\rightarrow$ <u>Properties</u> $\lambda_1, \lambda_2 \rightarrow$ eigenvalues of A , $\lambda_1 \neq \lambda_2$
  - ① $\lambda_1, \lambda_2 \in \mathbb{R}$
  - ② $Ax = \lambda_1 x$, $Ay = \lambda_2 y$, then $x \cdot y = \bar{x}^T \cdot y = 0$ i.e., $x \perp y$
  - ③ If no eigenvalues are repeated, then A is diagonalisable

③ Unitary Matrices

- $\rightarrow$ $U^* U = I$
- $\rightarrow$ orthonormal columns
- $\rightarrow$ <u>Properties</u> $\lambda_1, \lambda_2 \rightarrow$ eigenvalues of U
  - ① $\|Ux\| = \|x\|$
  - ② $|\lambda_1| = |\lambda_2| = 1$
  - ③ $Ux = \lambda_1$, $Uy = \lambda_2 y$, then $x \cdot y = \bar{x}^T \cdot y = 0$

④ Any $n \times n$ matrix is similar to an upper triangle matrix

- $\rightarrow$ $A = \underset{\text{unitary matrix}}{U} \; \underset{\text{upper triangle matrix}}{T} \; U^*$
- $\rightarrow$ <u>Procedure</u> :
  - ① $A = \begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}_{3 \times 3}$. Find eigenvector z_1 corresponding to λ_1
  - ② Extend z_1 to form orthonormal basis for $\mathbb{R}^3$. $\{z_1, e_1, e_2\}$
    - $\rightarrow$ Use Gram-Schmidt process on $\{z_1, e_1, e_2\} \rightarrow \{w_1, w_2, w_3\}$
  - ③ $U_1 = \begin{bmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{bmatrix}$. $U_1^* A U_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & \cdots & \cdots \\ 0 & \cdots & \cdots \end{bmatrix}$ B
  - ④ $B = \begin{bmatrix} \cdots & \cdots \\ \cdots & \cdots \end{bmatrix}$. Find eigenvector z_2 corresponding to λ_2 .

    Repeat ①-③ for B.

⑤ Single Value Decomposition

- $\rightarrow$ Any real $m \times n$ matrix A decomposed is SVD form:

  $\underset{m \times n}{A} = \underset{m \times m}{Q_1} \, \Sigma \, \underset{n \times n}{Q_2^T}$, where $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$, $D = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix}$

- $\rightarrow$ <u>Procedure</u> :
  - ① Find eigenvalues and eigenvectors of $A^T A$. $\sigma_i = \sqrt{\lambda_i}$ $\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{bmatrix}$
  - ② $\{x_1, \dots, x_n\} \rightarrow$ orthonormal eigenvectors of $A^T A$. $Q_2 = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}$
  - ③ $y_i = \frac{1}{\sigma_i} A x_i$. $Q_1 = \begin{bmatrix} | & & | \\ y_1 & \cdots & y_n \\ | & & | \end{bmatrix}$

⑥ Positive Definite

- $\rightarrow$ f is +ve definite if $f'(0) = 0$ and $f(x) > 0 \;\forall x$
- $\rightarrow$ A is +ve definite matrix if

  $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ $a > 0$ and $ac - b^2 > 0$

7. Principal Component Analysis
   - → dimensionality reduction
   - → <u>Procedure</u>:
     1. Data : $\{x_1, \dots, x_n\}$ $x_i \in \mathbb{R}^d$ $\forall i$. To reduce it to m-dimensions
     2. Let $\bar{x} = 1/n \sum_{i=1}^{n} x_i$ and $C = 1/n \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$
     3. Find eigenvalues $\{\lambda_1, \dots, \lambda_d\}$ $\lambda_1 \geq \dots \geq \lambda_d$ and eigenvectors $\{u_1, \dots, u_d\}$
     4. Project data $\{x_1, \dots, x_n\}$ using $\tilde{x}_i = \sum_{j=1}^{m} (x_i^T u_j) u_j + \sum_{j=m+1}^{d} (\bar{x}^T u_j) u_j$
     5. Reconstruction error : $J^* = 1/n \sum_{i=1}^{n} \| x_i - \tilde{x}_i \|^2$
     6. Projected variance : $1/n \sum_{i=1}^{n} (x_i^T u - \bar{x}^T u)^2$
        ↪ is equal to the eigenvalue of that PC

WEEKS 9-12

8. For x^* to be the optimal solution:
   $\nabla f(x^*) = -\lambda \nabla g(x^*)$ where $f(x) \rightarrow$ objective function and $g(x) \leq 0$

9. Properties of a convex function:
   1. Sum of convex functions is a convex function
   2. Composition of convex function with convex + non-decreasing is a convex function.
   3. Composition of linear with convex is a convex function.

10. If f, h are convex , then strong-duality holds
    i.e., $\underbrace{\min_x \left[\max_{\lambda \geq 0} L(x, \lambda) \right]}_{\text{Primal Problem}}$ (or) $\underbrace{\max_{\lambda \geq 0} \left[\min_x L(x, \lambda) \right]}_{\text{Dual Problem}}$

11. K.K.T. conditions:
    1. Stationarity : $\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0$
    2. Complimentary slackness : $\lambda^* h(x^*) = 0$
    3. Primal feasibility : $h(x^*) \leq 0$
    4. Dual feasibility : $\lambda^* \geq 0$

12. $X, Y \longrightarrow f_{XY}(x, y)$

    W, Z $\quad X = g(W, Z) \quad Y = h(W, Z)$

    $$f_{WZ} = f_{XY}(g(w,z), h(w,z)) \, |\text{Det } J| \text{ , where } J = \begin{bmatrix} \frac{\partial g}{\partial w} & \frac{\partial g}{\partial z} \\ \frac{\partial h}{\partial w} & \frac{\partial h}{\partial z} \end{bmatrix}$$

13. Gradient descent for linear reg → $w_{t+1} = w_t - \eta \nabla f(w_t)$, where $\nabla f(w_t) = (X^T X) w_t - X^T y$
