Intro to ML

Objective Function for Vanilla Linear Regression:

$$J(\omega) = \frac{1}{2}\|t - X\omega\|_2^2 \tag{1}$$

We solve for the optimal ω by taking the derivative of the objective function with respect to ω and setting it to zero:

$$\frac{\partial J(\omega)}{\partial \omega} = 0, (X^T X)^{-1} X^T t = \omega \tag{2}$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} \tag{3}$$

$$(X^T X + \lambda I)^{-1} X^T t = \omega \tag{4}$$

Experimental Design and Analysis

Basis Functions

$$\phi(x) = \begin{bmatrix} \phi_0(x) \\ \phi_1(x) \\ \vdots \\ \phi_M(x) \end{bmatrix} \tag{5}$$

$$\phi_j(x) = \exp - \frac{\|x - \mu\|^2}{2\sigma^2} \tag{6}$$

$$\phi_j(x) = x^j \tag{7}$$

Model Selection

$$\min_{\omega} \frac{1}{2}\|t - X\omega\|_2^2 + \lambda\|\omega\|_1 \tag{8}$$

$$\min_{\omega} \frac{1}{2}\|t - X\omega\|_2^2 + \lambda_1\|\omega\|_1 + \lambda_2\|\omega\|_2^2, \quad \lambda_1 + \lambda_2 = 1 \tag{9}$$

Metrics of Regression

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(t_i - \hat{t}_i)^2 \tag{10}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|t_i - \hat{t}_i| \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(t_i - \hat{t}_i)^2}{\sum_{i=1}^{n}(t_i - \bar{t})^2}, \quad \bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i \tag{12}$$

$$R^2 = 1 - \frac{\|t - X\omega\|_2^2}{\|t - \bar{t}\|_2^2} \tag{13}$$

Bayesian Learning

$$P(C_i|x) = \frac{P(x|C_i)P(C_i)}{\sum_{j=1}^{K} P(x|C_j)P(C_j)}$$

$$P(\lambda) = \frac{\beta_\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$

1. Gaussian-Gaussian
2. Gaussian-Exponential
3. Gaussian-Gamma
4. Gaussian-Beta
5. Gaussian-Dirichlet
6. Gaussian-Wishart
7. Gaussian-Inverse Wishart
8. Gaussian-Student's t
9. Gaussian-Laplace
10. Gaussian-Cauchy

Generative Models

$$p(t|x, \omega) = \mathcal{N}(t; \omega^T\phi(x), \beta^{-1}) \tag{14}$$

$$p(x|\omega) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Sigma_k)$$

$$\Theta = \{\pi_1, \pi_2, \ldots, \pi_K, \mu_1, \mu_2, \ldots, \mu_K, \Sigma_1, \Sigma_2, \ldots, \Sigma_K\}, \quad \sum_{k=1}^{K} \pi_k = 1$$

$$\mathcal{L}_0 = \prod_{i=1}^{N}\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)$$

$$\ln \mathcal{L}_0 = \sum_{i=1}^{N} \ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)$$

z_i = label of the Gaussian component for the i^{th} data point x_i

$$\mathcal{L}^c = \prod_{i=1}^{N} \pi_{z_i} \mathcal{N}(x_i; \mu_{z_i}, \Sigma_{z_i})$$

$$Q(\Theta, \Theta^{(t)}) = \mathbb{E}_z[\ln p(x, z|\Theta)|X, \Theta^{(t)}]$$

$$P(z_i|x_i, \Theta^{(t)}) = \frac{P(x_i|z_i, \Theta^t)P(z_i|\Theta^t)}{P(x_i|\Theta^t)} = \frac{\pi_{z_i}\mathcal{N}(x_i; \mu_{z_i}, \Sigma_{z_i})}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}$$

$$C_{ik} = P(z_i|x_i, \Theta^{(t)}) = \frac{\pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i; \mu_j, \Sigma_j)}$$

$$\arg_\Theta \max Q(\Theta, \Theta^{(t)}) = \sum_{z_i}^{K} \ln(\mathcal{L}^c) P(z_i|x_i, \Theta^{(t)})$$

$$= \sum_{z_i=1}^{K} \ln\left(\prod_{i=1}^{N} \pi_{z_i} G(x_i; \mu_{z_i}, \Sigma_{z_i})\right) P(z_i|x_i, \Theta^{(t)})$$

$$= \sum_{k=1}^{K}\sum_{i=1}^{N} (\ln(\pi_k) + \ln(G(x_i; \mu_k, \Sigma_k))) C_{ik}$$

$$= \sum_{k=1}^{K}\sum_{i=1}^{N}(\ln(\pi_k) - \frac{d}{2}\ln(2\pi) - \frac{d}{2}\ln(\sigma_k^2) - \frac{1}{2\sigma_k^2}\|x_i - \mu_k\|_2^2)C_{ik}$$

$$\mu_K = \frac{\sum_{i=1}^{N}\mathbf{x}_i C_{ik}}{\sum_{i=1}^{N} C_{ik}}$$

$$\sigma_k^2 = \frac{\sum_{i=1}^{N} C_{ik}\|x_i - \mu_k\|_2^2}{d\sum_{i=1}^{N} C_{ik}}$$

$$\pi_k = \frac{\sum_{i=1}^{N} C_{ik}}{N}$$

Non-Parametric Models

$$J(\Theta, U) = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik} d^2(x_i, \theta_k) = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}\|x_i - \theta_k\|_2^2$$

$$\theta_k = \frac{\sum_{x_i \in C_k} x_i}{N_k}$$

1. b_i = average distance from x_i to all other points in the same cluster
2. a_i = average distance from x_i to all other points in the nearest cluster

$$\text{Silhouette} = \frac{1}{N}\sum_{i=1}^{N}\frac{b_i - a_i}{\max(a_i, b_i)}$$

1. a = number of pairs of elements in S that are in the same cluster and in the same set in S'
2. b = number of pairs of elements in S that are in different clusters and in different sets in S'
3. c = number of pairs of elements in S that are in the same cluster and in different sets in S'
4. d = number of pairs of elements in S that are in different clusters and in the same set in S'

$$\text{Rand Index} = \frac{a + b}{a + b + c + d}$$

$$\text{Jaccard} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}$$

$$\text{Euclidean Distance} = \|x_1 - x_2\|_2^2 = \sum_{i=1}^{d}(x_{1i} - x_{2i})^2$$

$$\text{Manhattan Distance} = \|x_1 - x_2\|_1 = \sum_{i=1}^{d}|x_{1i} - x_{2i}|$$

$$\text{Mahalanobis Distance} = \|x_1 - x_2\|_\Sigma = \sqrt{(x_1 - x_2)^T\Sigma^{-1}(x_1 - x_2)}$$