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

$$\omega_{MLE} = \text{arg}_\omega \text{max} \prod_{i=1}^{n} p(t_i|x_i, \omega) \tag{14}$$

$$\omega_{MLE} = \text{arg}_\omega \text{max} \sum_{i=1}^{N} \ln p(t_i|x_i, \omega) \tag{15}$$

$$\omega_{MLE} = \text{arg}_\omega \text{min} \mathbb{E}_{\phi(x)}[\ln p(t|x, \omega)]$$

$$\omega_{\text{MAP}} \propto \text{arg}_\omega \text{max} \prod_{i=1}^{N} \mathcal{N}(t_i; y_i, 1)\mathcal{N}(\omega_j; 0, \frac{1}{\lambda}) \tag{16}$$

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

$$p(t|x, \omega) = \mathcal{N}(t; \omega^T \phi(x), \beta^{-1}) \tag{17}$$

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

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

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

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

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