BIOENG-210: Biological Data Science I: Statistical Learning

Required Formulas

Lecture 1

Expectation

$$\mathbb{E}[X] = \sum_x x \cdot \mathbb{P}(X = x) \text{ or } \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx$$

Variance

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Normal Distribution PDF

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Poisson Distribution

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

Lecture 2

Maximum Likelihood Estimation

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta}\ L(\theta \mid x_1, x_2, \ldots, x_n) = \arg\max_{\theta}\ \prod_{i=1}^{n} f(x_i \mid \theta)$$

MLE equivalent minimize the negative log-likelihood

$$\hat{\theta}_{\mathrm{MLE}} = \arg\min_{\theta}\ -\log L(\theta \mid x_1, x_2, \ldots, x_n) = \arg\min_{\theta}\ -\sum_{i=1}^{n} \log f(x_i \mid \theta)$$

Covariance

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Correlation

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Pearson Correlation

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Lecture 3

Joint Probability Distribution

$$P((X,Y)\in A)=\iint_A f_{X,Y}(x,y)\,dx\,dy$$

Conditional PDF and Chain Rule

$$f_{Y|X}(y|x)=\frac{f_{X,Y}(x,y)}{f_X(x)}$$

Mutual Information

$$I(X;Y)=\iint f_{X,Y}(x,y)\,\log\frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)}\,dx\,dy$$

Lecture 4

Sample Mean and Its Properties

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad \mathbb{E}[\bar{X}]=\mu,\quad \mathrm{Var}(\bar{X})=\frac{\sigma^2}{n}$$

Central Limit Theorem

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\xrightarrow{d}\mathcal{N}(0,1)$$

One-Sample t-Test Statistic

$$t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\quad\sim\ t_{n-1}$$

Benjamini–Hochberg (BH) Threshold

$$p_{(k)}\le\frac{k}{m}\,\alpha$$

Lecture 5

Least Squares Estimators

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2};\quad \hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$$

Residual Sum of Squares (RSS)

$$\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1x_i))^2=\mathrm{RSS}(\beta_0,\beta_1)$$

Sampling Distribution of Slope Estimator

$$\hat{\beta}_1\sim\mathcal{N}\left(\beta_1,\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

Confidence Interval at $100(1-\alpha)\%$ for regression parameters

$$\hat{\beta}\pm t_{\alpha/2,n-2}\times SE(\hat{\beta})$$

Lecture 6

Multiple Regression Model (Matrix Form)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Normal Equations

$$\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y}$$

Solution to Normal Equations (if X has full column rank)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

RSS for Multiple Linear Regression

$$\text{RSS}(\boldsymbol{\beta}) = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Coefficient of Determination R^2

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Lecture 8

Poisson GLM

Random component: $Y \sim \text{Poisson}(\lambda)$

Link function: $\eta = \log(E[Y]) = \log(\lambda)$

Inverse link function: $E[Y] = \lambda = e^\eta$

Logistic regression

Random component: $Y \sim \text{Bernoulli}(p)$

Link function: $\eta = g(p) = \log\left(\frac{p}{1-p}\right)$

Inverse link function: $p = g^{-1}(\eta) = \frac{e^\eta}{1+e^\eta}$

Classification metrics

$$\textbf{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} = \frac{\text{Correct predictions}}{\text{Total samples}}$$

$$\textbf{Sensitivity/Recall} = \frac{TP}{TP + FN} = \frac{\text{Correctly predicted positive}}{\text{Total positive samples}}$$

$$\textbf{Precision} = \frac{TP}{TP + FP} = \frac{\text{Correctly predicted positive}}{\text{Total predicted positive}}$$

Lecture 9

Ridge regression

Objective: $\min_\beta \lVert y - X\beta \rVert_2^2 + \lambda \lVert \beta \rVert_2^2$

Bayes' Theorem

$$P(\theta \mid y) = \frac{P(y \mid \theta)\, P(\theta)}{P(y)}$$

Maximum A Posteriori (MAP) Estimation

$$\begin{aligned}\hat{\theta}_{\text{MAP}} &= \arg\max_\theta P(\theta \mid y) \\ &= \arg\max_\theta \left[P(y \mid \theta)\, P(\theta)\right]\end{aligned}$$

MSE Definition

$$\text{MSE}\big(\hat{f}(x)\big) = \mathbb{E}\big[(Y - \hat{f}(x))^2\big]$$

Bias (squared)

$$\text{Bias}^2\big(\hat{f}(x)\big) = (f(x) - \mathbb{E}[\hat{f}(x)])^2$$

Variance

$$\text{Var}\big(\hat{f}(x)\big) = \mathbb{E}[(\mathbb{E}[\hat{f}(x)] - \hat{f}(x))^2]$$

Bias-Variance Decomposition

$$\text{MSE}\big(\hat{f}(x)\big) = \text{Bias}^2\big(\hat{f}(x)\big) + \text{Var}\big(\hat{f}(x)\big) + \sigma_\epsilon^2$$

Lecture 10

Multivariate Normal

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right); \quad \boldsymbol{\mu} = E[\mathbf{X}]; \quad \boldsymbol{\Sigma} = E[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T]$$

Singular value decomposition (SVD)

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

SVD properties

$$\mathbf{X} \in \mathbb{R}^{n\times p} \quad \mathbf{U} \in \mathbb{R}^{n\times n}, \mathbf{U}^T\mathbf{U} = \mathbf{I_n} \quad \mathbf{V} \in \mathbb{R}^{p\times p}, \mathbf{V}^T\mathbf{V} = \mathbf{I_p} \quad \mathbf{S} \in \mathbb{R}^{n\times p}(\text{Rect. diag})$$

Lecture 11

Proportion of variance explained in PCA

$$\text{Proportion of variance explained} = \frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{\min(n,p)} \sigma_i^2} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i}$$

Single Linkage

$$d(C_i, C_j) = \min_{\mathbf{x}\in C_i, \mathbf{y}\in C_j} d(\mathbf{x}, \mathbf{y})$$

Average Linkage

$$d(C_i, C_j) = \sum_{\mathbf{x}\in C_i} \sum_{\mathbf{y}\in C_j} \frac{d(\mathbf{x}, \mathbf{y})}{|C_i||C_j|}$$

Lecture 12

No formulas to memorize for this lecture!