

Midterm 1 note sheet

Linear regression

Intercept trick

$$D(\tilde{x}) = [1, x_1, \dots, x_p]^\top$$

Predictor matrix

$$\mathbf{x} = [\tilde{x}_1, \dots, \tilde{x}_n]^\top \in \mathbb{R}^{n \times p}$$

Design matrix

$$D(\mathbf{x}) = [D(\tilde{x}_1), \dots, D(\tilde{x}_n)]^\top$$

Design matrix shape

$$D(\mathbf{x}) \in \mathbb{R}^{n \times p^*}$$

Coefficient vector shape

$$\vec{\beta} \in \mathbb{R}^{p^*}$$

Predictions

$$\hat{y}_{\tilde{x}} = D(\mathbf{x})\vec{\beta}$$

Risk

$$R(\vec{\beta}) = \mathbb{E}[(\vec{\beta} \cdot D(\tilde{x}) - y)^2]$$

Sample MSE

$$\text{sMSE}(\vec{\beta}) = \frac{1}{n} \sum_{i=1}^n (\vec{\beta} \cdot D(\tilde{x}_i) - y_i)^2$$

Sample MSE, matrix form

$$\text{sMSE}(\vec{\beta}) = \frac{1}{n} \|D(\mathbf{x})\vec{\beta} - y\|_2^2$$

Normal equations

$$D(\mathbf{x})^\top D(\mathbf{x})\hat{\vec{\beta}} = D(\mathbf{x})^\top y_{\tilde{x}}$$

Unique solution condition

$$\det(D(\mathbf{x})^\top D(\mathbf{x})) \neq 0$$

Solution

$$\hat{\vec{\beta}} = (D(\mathbf{x})^\top D(\mathbf{x}))^{-1} D(\mathbf{x})^\top y_{\tilde{x}}$$

Fitted values

$$\hat{y}_{\tilde{x}} = D(\mathbf{x})\hat{\vec{\beta}}$$

Residual vector

$$\hat{\epsilon}_{\tilde{x}} = y_{\tilde{x}} - \hat{y}_{\tilde{x}}$$

Feature expansion

Feature map

$$D : \mathbb{R}^p \rightarrow \mathbb{R}^{p^*}$$

Linear model in the features

$$h(\tilde{x}) = \vec{\beta} \cdot D(\tilde{x})$$

Monomial

$$m(\tilde{x}) = \prod_{j=1}^p x_j^{a_j}$$

Monomial degree bound

$$\sum_{j=1}^p a_j \leq d$$

Shapes and calculus

Matrix product shape

$$(a \times b)(b \times c) \rightarrow a \times c$$

Transpose shape

$$a \times b \rightarrow b \times a$$

Dot product

$$\tilde{x} \cdot \tilde{y} = \sum_{i=1}^n x_i y_i$$

Sample mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Sum rule

$$\frac{d}{dw} \sum_{i=1}^n u_i = \sum_{i=1}^n \frac{du_i}{dw}$$

Chain rule

$$\frac{dh(u)}{dw} = \frac{dh(u)}{du} \frac{du}{dw}$$

Minimum: first derivative is zero

$$\frac{\partial \hat{R}}{\partial w} = 0$$

Minimum: second derivative is positive

$$\frac{\partial^2 \hat{R}}{\partial w^2} > 0$$

Vector calculus

Linear form	$\frac{\partial}{\partial \vec{\beta}} \mathcal{L} \cdot \vec{\beta} = \mathcal{L}$
Squared norm	$\frac{\partial}{\partial \vec{\beta}} \vec{\beta} \cdot \vec{\beta} = 2\vec{\beta}$
Matrix-vector product, constant $\mathbf{A}$	$\frac{\partial}{\partial \vec{\beta}} (\mathbf{A}v) = (\frac{\partial}{\partial \vec{\beta}} v) \mathbf{A}^\top$
Special case $v = \vec{\beta}$	$\frac{\partial}{\partial \vec{\beta}} (\mathbf{A}\vec{\beta}) = \mathbf{A}^\top$
Next rule needs $\mathbf{S}$ symmetric	$\mathbf{S} = \mathbf{S}^\top$
Quadratic form	$\frac{\partial}{\partial \vec{\beta}} \vec{\beta}^\top \mathbf{S} \vec{\beta} = 2\mathbf{S}\vec{\beta}$

Generalization

Outcome	$y = \mu(x) + \varepsilon$
Noise variance	$\text{Var}(\varepsilon) = \sigma^2$
Bias	$\text{Bias}(\hat{f}(x)) = \mathbb{E}[\hat{f}(x)] - \mu(x)$
Variance	$\text{Var}(\hat{f}(x)) = \mathbb{E}[(\hat{f}(x) - \mathbb{E}[\hat{f}(x)])^2]$
Expected squared error	$\text{ESE}(x) = \mathbb{E}[(y - \hat{f}(x))^2]$
Expected squared error decomposition	$\text{ESE}(x) = (\text{Bias}(\hat{f}(x)))^2 + \text{Var}(\hat{f}(x)) + \sigma^2$
Cross-validation estimate, equal-size-fold case	$\hat{R}_{k\text{-fold}} = \frac{1}{k} \sum_{j=1}^k \text{MSE}_j$

	e	e^{-1}	$\ln 2$	$\ln 3$	$\ln 4$	$\ln 5$	$\ln 10$
Value	2.718	0.368	0.693	1.099	1.386	1.609	2.303
