Model selection and averaging

David J. H. Shih

Intended learning outcomes

- Apply model selection and model averaging to statistical models.

Predictive models

Use data $(\boldsymbol{x}, \boldsymbol{y})$ to estimate parameters $\boldsymbol{\theta}$ of the model and to predict new outcome $\hat{y}$ given new data point $\tilde{x}$.

Discriminative model with MLE

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p\left(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}\right)$$

$$\hat{y} = \underset{y}{\operatorname{argmax}}\, p\left(y \mid \tilde{x}, \hat{\boldsymbol{\theta}}\right)$$

Discriminative model with MAP

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p\left(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{x}\right) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p\left(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}\right)\, p\left(\boldsymbol{\theta}\right)$$

$$\hat{y} = \underset{y}{\operatorname{argmax}}\, p\left(y \mid \tilde{x}, \hat{\boldsymbol{\theta}}\right)$$

Generative model with MLE

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} p\left(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}\right)$$

$$\hat{y} = \operatorname*{argmax}_{y} p\left(\tilde{x} \mid \hat{\boldsymbol{\theta}}, y\right)$$

Generative model with MAP

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} p\left(\boldsymbol{\theta} \mid \boldsymbol{x}, \boldsymbol{y}\right) = \operatorname*{argmax}_{\boldsymbol{\theta}} p\left(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}\right)\, p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)$$

$$\hat{y} = \operatorname*{argmax}_{y} p\left(y \mid \tilde{x}, \hat{\boldsymbol{\theta}}\right) = \operatorname*{argmax}_{y} p\left(\tilde{x} \mid \hat{\boldsymbol{\theta}}, y\right)\, p\left(y\right)$$

Full Bayesian discriminative model

$$p\left(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{x}\right) = \frac{p\left(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}\right)\, p\left(\boldsymbol{\theta}\right)}{p\left(\boldsymbol{y} \mid \boldsymbol{x}\right)}$$

$$p\left(\tilde{y} \mid \tilde{x}\right) = \int_{\boldsymbol{\Theta}} p\left(\tilde{y} \mid \tilde{x}, \boldsymbol{\theta}\right)\, p\left(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{x}\right) d\boldsymbol{\theta}$$

Full Bayesian generative model

$$p\left(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{x}\right) = \frac{p\left(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}\right)\, p\left(\boldsymbol{\theta} \mid \boldsymbol{y}\right)}{p\left(\boldsymbol{x} \mid \boldsymbol{y}\right)}$$

$$p\left(\tilde{x} \mid \tilde{y}\right) = \int_{\boldsymbol{\Theta}} p\left(\tilde{x} \mid \boldsymbol{\theta}, \tilde{y}\right)\, p\left(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{x}\right) d\boldsymbol{\theta}$$

$$p\left(\tilde{y} \mid \tilde{x}\right) = \frac{p\left(\tilde{x} \mid \tilde{y}\right)\, p\left(\tilde{y}\right)}{p\left(\tilde{x}\right)}$$

Summary

"All Models Are Wrong, Some are Useful." - George Box

Intended learning outcomes

- Apply model selection and model averaging to statistical models.