# STAT 543/641 – WINTER 2019 – HOMEWORK #2

DUE MARCH ??, 2019

---

Let $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ be random samples from populations with means $\mu_X$ and $\mu_Y$ and variances $\sigma_X^2$ and $\sigma_Y^2$, respectively. Let $S_X^2$ and $S_Y^2$ be the standard unbiased estimators of $\sigma_X^2$ and $\sigma_Y^2$, respectively.

(1) Suppose $\sigma_X^2 = \sigma_Y^2$ and write $\sigma^2$ for this common value.

$$S := \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$$

is an unbiased estimator of $\sigma^2$. It's called the *pooled variance estimator*.

(2) Suppose, in addition to having common variance, that the $X_i$ are independent of the $Y_i$. What is the distribution of $S_X^2$? What is its variance? Compare the mean squared errors $S_X^2$, $S_Y^2$, and $S^2$.

(3) Generalize these results from the case of $K = 2$ populations to that of an arbitrary $K$. Compare with equation (4.15) in [2].

(4) [**Bonus**] Can you prove analogous results with covariance matrices in place of variances?

---

[1, Exercise 12.16] This exercise examines an extreme case in which the likelihood equations for logistic regression have no solution.

Consider the following 20-point data set:

$$(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1),(0,1)$$
$$(1,0),(1,0),(1,0),(1,0),(1,0),(1,1),(1,1),(1,1),(1,1),(1,1)$$

(1) Observe that, empirically, $\mathrm{Prob}(Y = 1|X = 0) = 1$ and $\mathrm{Prob}(Y = 1|X = 1) = 0.5$. Let $\sigma(t) = (1 + e^{-t})^{-1}$ be the sigmoid function. Are there $a$ and $b$ such that $\sigma(a + b \cdot 0) = 1$ and $\sigma(a + b \cdot 1) = 0.5$?

(2) Let $\mathscr{L}(a, b)$ be the likelihood function associated to fitting a logistic regression model to this data set. Show that

$$L := \lim_{b\to\infty} \mathscr{L}(-b, b) = \sup_{(a,b)\in\mathbb{R}^2} \mathscr{L}(a, b) < \infty$$

and that $\mathscr{L}(a, b) \neq L$ for any $(a, b) \in \mathbb{R}^2$. What are

$$\lim_{b\to\infty} \sigma(-b + b \cdot 0) \quad \text{and} \quad \lim_{b\to\infty} \sigma(-b + b \cdot 1)?$$

Let $(\boldsymbol{X}, Y)$ be jointly distributed, where $\boldsymbol{X}$ is a $p$-dimensional random vector and $Y$ takes values in $\{1, \ldots, K\}$. Suppose that, for each $k$, $\boldsymbol{X}|Y = k$ has Gaussian distribution with mean $\boldsymbol{\mu}_k$ and and variance $\Sigma$, with the latter independent of $k$.

Consider a data set $(\boldsymbol{x}^{(1)}, y^{(1)}), \ldots, (\boldsymbol{x}^{(n)}, y^{(n)})$, where $\boldsymbol{x}^{(i)} \in \mathbb{R}^{1\times p}$ and $y^{(i)} \in \{1, \ldots, K\}$. For $1 \leq k \leq K$, let

$$I_k = \{i : y^{(i)} = k\}, \quad n_k = |I_k|, \quad \widehat{\pi}_k = \frac{n_k}{n}.$$

Define sample means $\boldsymbol{\mu}_k$ and a pooled sample covariance $\boldsymbol{\Sigma}$ by

$$\widehat{\boldsymbol{\mu}}_k = \widehat{\boldsymbol{\mu}}_{k,\boldsymbol{x}} = \frac{1}{n_k} \sum_{i\in I_k} \boldsymbol{x}^{(i)} \in \mathbb{R}^{p\times 1},$$

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} = \frac{1}{n-K} \sum_{k=1}^{K} \sum_{i\in I_k} (\boldsymbol{x}^{(i)} - \widehat{\boldsymbol{\mu}}_k)^T (\boldsymbol{x}^{(i)} - \widehat{\boldsymbol{\mu}}_k) \in \mathbb{R}^{p\times p}.$$

Define linear discriminant functions, $\delta_k = \delta_{k,\boldsymbol{x}}$, by

$$\delta_k(\boldsymbol{v}) = \delta_{k,\boldsymbol{x}}(\boldsymbol{v}) = \boldsymbol{v}\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\mu}}_k^T - \frac{1}{2}\widehat{\boldsymbol{\mu}}_k\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\mu}}_k^T + \log\widehat{\pi}_k, \quad \boldsymbol{v} \in \mathbb{R}^{p\times 1}.$$

Let $\boldsymbol{a} \in \mathbb{R}^{p\times 1}$ and let

$$\boldsymbol{w}^{(i)} = \boldsymbol{x}^{(i)} - \boldsymbol{a}.$$

$$\widehat{\boldsymbol{\mu}}_{k,\boldsymbol{w}} = \widehat{\boldsymbol{\mu}}_{k,\boldsymbol{x}} - \boldsymbol{a}, \quad \Sigma_w = \Sigma_x$$

$$\delta_{k_1,w}(v-a) - \delta_{k_2,w}(v-a) = \delta_{k_1,x}(v) - \delta_{k_2,x}(v)$$

Let $U \in \mathbb{R}^{p\times p}$ be an orthogonal matrix and let $w^{(i)} = Ux^{(i)}$. Then

$$\delta_{k,Ux}(Uv) = \delta_x(v).$$

$$\sum_{k=1}^{K} \pi_k \mu_k = \mu$$

Logistic regression (with and without ridge regularization, with and without PCA), LDA, Gaussian naïve Bayes, for breast cancer data set. Plot in 2d with decision boundary. Optional: Lasso

Document classification with multinomial naïve Bayes

Ridge regression via constrained optimization.

## References


[1] Casella, Bergger, *Statistical Inference (2nd ed.)*, Duxbury, 2002.

[2]