

Mid-term solutions
ECE 271A
Electrical and Computer Engineering
University of California San Diego

Nuno Vasconcelos

Fall 2018

1. a) The log-likelihood is

$$l(\theta) = \sum_i \log h(x_i) + \nu(\theta) \sum_i T(x_i) - nA(\theta).$$

Setting its derivative to zero, we have

$$\frac{\partial \nu}{\partial \theta} \sum_i T(x_i) = n \frac{\partial A}{\partial \theta},$$

and

$$\frac{1}{n} \sum_i T(x_i) = \frac{\partial A}{\partial \theta} \frac{1}{\frac{\partial \nu}{\partial \theta}}.$$

For this solution to be a maximum we need that

$$\frac{\partial^2 l}{\partial \theta^2} = \frac{\partial^2 \nu}{\partial \theta^2} \sum_i T(x_i) - n \frac{\partial^2 A}{\partial \theta^2} \leq 0$$

i.e. the additional constraint that

$$\frac{\partial^2 A}{\partial \theta^2} \geq \frac{\partial^2 \nu}{\partial \theta^2} \frac{1}{n} \sum_i T(x_i)$$

Plugging in the value of $\frac{1}{n} \sum_i T(x_i)$ at the critical point leads to the condition

$$\frac{\partial^2 A}{\partial \theta^2} \geq \frac{\partial^2 \nu}{\partial \theta^2} \frac{\partial A}{\partial \theta} \frac{1}{\frac{\partial \nu}{\partial \theta}}.$$

For the Gaussian, $\theta = \mu$, $T(x) = x$ and the derivatives are

$$\begin{aligned} \frac{\partial A}{\partial \mu} &= \mu \\ \frac{\partial^2 A}{\partial \mu^2} &= 1 \\ \frac{\partial \nu}{\partial \mu} &= 1 \\ \frac{\partial^2 \nu}{\partial \mu^2} &= 0. \end{aligned}$$

The ML estimate is

$$\mu = \frac{1}{n} \sum_i x_i$$

and the condition is

$$1 \geq 0$$

which always holds. Hence, we have a maximum. For the binomial, $\theta = p$, $T(x) = x$ and the derivatives are

$$\begin{aligned}\frac{\partial A}{\partial p} &= \frac{\partial}{\partial p}(-n \log(1-p)) = \frac{n}{1-p} \\ \frac{\partial^2 A}{\partial p^2} &= \frac{n}{(1-p)^2} \\ \frac{\partial \nu}{\partial p} &= \frac{\partial}{\partial p}(\log p - \log(1-p)) = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)} \\ \frac{\partial^2 \nu}{\partial p^2} &= -\frac{1-2p}{p^2(1-p)^2}\end{aligned}$$

The ML estimate is

$$\begin{aligned}np &= \frac{1}{n} \sum_i x_i \\ p &= \frac{1}{n^2} \sum_i x_i\end{aligned}$$

and the condition

$$\begin{aligned}\frac{\partial^2 A}{\partial p^2} &\geq \frac{\partial^2 \nu}{\partial p^2} \frac{\partial A}{\partial p} \frac{1}{\frac{\partial \nu}{\partial p}} \\ \frac{n}{(1-p)^2} &\geq -\frac{1-2p}{p^2(1-p)^2} \frac{n}{1-p} p(1-p) \\ 1 &\geq -\frac{1-2p}{p} \\ p &\geq 2p-1 \\ 1 &\geq p\end{aligned}$$

always holds. Hence, we have a maximum.

b) According to bayes decision rule, the optimal decision function, under the “0/1” loss is “pick $i = 0$ ” if

$$\begin{aligned}P_{Y|X}(0|\mathcal{D}) &> P_{Y|X}(1|\mathcal{D}) \\ P_{X|Y}(\mathcal{D}|0)P_Y(0) &> P_{X|Y}(\mathcal{D}|1)P_Y(1) \\ P_{X|Y}(x_1, \dots, x_n|0)P_Y(0) &> P_{X|Y}(x_1, \dots, x_n|1)P_Y(1)\end{aligned}$$

Taking the log on both sides, and using the independence assumption,

$$\begin{aligned}\log \frac{P_{X|Y}(x_1, \dots, x_n|0)}{P_{X|Y}(x_1, \dots, x_n|1)} + \log \frac{P_Y(0)}{P_Y(1)} &> 0 \\ [\nu(\theta_0) - \nu(\theta_1)] \sum_i T(x_i) - n[A(\theta_0) - A(\theta_1)] + \log \frac{\pi_0}{\pi_1} &> 0 \\ s_n &> \frac{1}{[\nu(\theta_0) - \nu(\theta_1)]} \left([A(\theta_0) - A(\theta_1)] + \frac{1}{n} \log \frac{\pi_1}{\pi_0} \right),\end{aligned}$$

where we have assumed that $\nu(\theta_0) \geq \nu(\theta_1)$. Therefore, the threshold is, $T = \frac{1}{[\nu(\theta_0) - \nu(\theta_1)]} \left([A(\theta_0) - A(\theta_1)] + \frac{1}{n} \log \frac{\pi_1}{\pi_0} \right)$.

2.a) The log-likelihood function is

$$\mathcal{L}(\theta) = \log(\theta) \sum_i x_i + \log(1 - \theta) \sum_i (1 - x_i)$$

and has zero derivative, with respect to θ , when

$$\begin{aligned} \frac{1}{\theta} \sum_i x_i &= \frac{1}{1 - \theta} \sum_i (1 - x_i) \\ \sum_i x_i - \theta \sum_i x_i &= n\theta - \theta \sum_i x_i \end{aligned}$$

from which

$$\theta_{ML} = \frac{1}{n} \sum_i x_i. \quad (1)$$

Also we check the second order derivative.

$$\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta^2} = -\frac{1}{\theta^2} \sum_i x_i - \frac{1}{(1 - \theta)^2} \sum_i (1 - x_i) \leq 0,$$

from which we have a maximum at θ_{ML} .

b) Solving the transformation for γ , we have

$$\begin{aligned} (1 - \theta)e^\gamma &= \theta \\ \theta(1 + e^\gamma) &= e^\gamma \\ \theta &= \frac{e^\gamma}{1 + e^\gamma} = \frac{1}{1 + e^{-\gamma}}. \end{aligned}$$

It follows that

$$\begin{aligned} P_X(x; \gamma) &= \left(\frac{e^\gamma}{1 + e^\gamma} \right)^x \left(\frac{1}{1 + e^\gamma} \right)^{1-x} \\ &= \frac{1}{1 + e^\gamma} e^{\gamma x}. \end{aligned}$$

c) The log-likelihood function is

$$\mathcal{L}(\gamma) = -n \log(1 + e^\gamma) + \gamma \sum_i x_i$$

and has zero derivative, with respect to γ , when

$$\begin{aligned} \sum_i x_i &= n \frac{e^\gamma}{1 + e^\gamma} \\ \frac{e^\gamma}{1 + e^\gamma} &= \frac{1}{n} \sum_i x_i = \theta_{ML} \end{aligned}$$

from which

$$\gamma_{ML} = \log \left(\frac{\theta_{ML}}{1 - \theta_{ML}} \right). \quad (2)$$

Also we check the second order derivative.

$$\frac{\partial^2 \mathcal{L}(\gamma)}{\partial \gamma^2} = -n \frac{e^\gamma}{(1 + e^\gamma)^2} \leq 0,$$

from which we have a maximum at γ_{ML} .

c) The log-likelihood function with respect to γ is

$$\mathcal{L}(\gamma) = \sum_i \log P_X(x_i; \gamma)$$

and has zero derivative when

$$\sum_i \frac{1}{P_X(x_i; \gamma)} \frac{d}{d\gamma} P_X(x_i; \gamma) = 0.$$

γ_{ML} is the solution of this equation. The log-likelihood function with respect to θ is

$$\mathcal{L}(\theta) = \sum_i \log P_X(x_i; f(\theta))$$

and has zero derivative when

$$\begin{aligned} 0 &= \sum_i \frac{1}{P_X(x_i; f(\theta))} \frac{d}{d\gamma} P_X(x_i; \gamma) \Big|_{\gamma=f(\theta)} \frac{d}{d\theta} f(\theta) \\ 0 &= \frac{d}{d\theta} f(\theta) \left[\sum_i \frac{1}{P_X(x_i; \gamma)} \frac{d}{d\gamma} P_X(x_i; \gamma) \right]_{\gamma=f(\theta)}. \end{aligned}$$

Since $f(\theta)$ is an invertible transformation, this holds if and only if

$$0 = \left[\sum_i \frac{1}{P_X(x_i; \gamma)} \frac{d}{d\gamma} P_X(x_i; \gamma) \right]_{\gamma=f(\theta)}$$

i.e. if and only if $f(\theta) = \gamma_{ML}$. Hence,

$$\gamma_{ML} = f(\theta_{ML}).$$

3. a) According to bayes decision rule, the optimal decision function, under the “0/1” loss is “pick 1 over 2” if

$$\begin{aligned} P_{Y|X}(1|x) &> P_{Y|X}(2|x) \\ P_{X|Y}(x|1)P_Y(1) &> P_{X|Y}(x|2)P_Y(2) \\ P_{X|Y}(x|1) &> P_{X|Y}(x|2) \text{ (as the priors are equal)} \end{aligned}$$

Taking Log on both sides,

$$\begin{aligned} \log P_{X|Y}(x|1) - \log P_{X|Y}(x|2) &> 0 \\ (x - \mu_2)^T \Sigma^{-1}(x - \mu_2) - (x - \mu_1)^T \Sigma^{-1}(x - \mu_1) &> 0 \\ 2(\mu_2 - \mu_2)^T \Sigma^{-1}x - (\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) &> 0 \\ \Rightarrow (\mu_1 - \mu_2)^T \Sigma^{-1} \left(x - \frac{1}{2}(\mu_1 + \mu_2) \right) &> 0 \end{aligned}$$

So the decision boundary can be expressed as a hyperplane, $w^T(x - x_0) = 0$, where w is the normal and x_0 is a point in the hyperplane, with

$$w = \Sigma^{-1}(\mu_1 - \mu_2) \quad (3)$$

$$x_0 = \frac{1}{2}(\mu_1 + \mu_2) \quad (4)$$

b) In this case

$$\begin{aligned} \hat{w} &= \Sigma^{-1}(\hat{\mu}_1 - \hat{\mu}_2) \\ \hat{x}_0 &= \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2) \end{aligned}$$

and

$$\begin{aligned} E_{X_1, \dots, X_{2n}}[\hat{w}] &= \Sigma^{-1}(E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2]) \\ E_{X_1, \dots, X_{2n}}[\hat{x}_0] &= \frac{1}{2}(E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] + E_{X_1, \dots, X_{2n}}[\hat{\mu}_2]). \end{aligned}$$

Hence, \hat{w} and \hat{x}_0 are both unbiased if and only if

$$\begin{cases} \Sigma^{-1}(E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2]) = \Sigma^{-1}(\mu_1 - \mu_2) \\ \frac{1}{2}(E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] + E_{X_1, \dots, X_{2n}}[\hat{\mu}_2]) = \frac{1}{2}(\mu_1 + \mu_2) \end{cases}$$

or

$$\begin{cases} E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2] = \mu_1 - \mu_2 \\ E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] = \mu_1 + \mu_2 - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2] \end{cases}$$

or

$$\begin{cases} E_{X_1, \dots, X_{2n}}[\hat{\mu}_2] = \mu_2 \\ E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] = \mu_1 \end{cases}$$

Hence, *both* $\hat{\mu}_i$ have to be unbiased.

c) In this case, we only need

$$\Sigma^{-1}(E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2]) = \Sigma^{-1}(\mu_1 - \mu_2).$$

Since Σ is invertible, this is equivalent to

$$E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2] = \mu_1 - \mu_2.$$

The $\hat{\mu}_i$ do not have to be unbiased, since \hat{w} will be unbiased as long the bias is a common constant, i.e.

$$\begin{aligned} E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] &= \mu_1 + \alpha \\ E_{X_1, \dots, X_{2n}}[\hat{\mu}_2] &= \mu_2 + \alpha \end{aligned}$$

where α is some non-zero vector.

d) In this case, we need

$$\Phi \Lambda \Phi^T (E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2]) = \Phi \Lambda \Phi^T (\mu_1 - \mu_2).$$

or, using the fact that Φ is invertible,

$$\Lambda \Phi^T (E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2]) = \Lambda \Phi^T (\mu_1 - \mu_2).$$

Let $\mathbf{a} = \Phi^T (E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2])$, $\mathbf{b} = \Phi^T (\mu_1 - \mu_2)$. Since the rows of Φ^T are the eigenvectors of Σ^{-1} , \mathbf{a} and \mathbf{b} are the vectors of coefficients of the projections of $(E_{X_1, \dots, X_{2n}}[\hat{\mu}_1] - E_{X_1, \dots, X_{2n}}[\hat{\mu}_2])$ and $(\mu_1 - \mu_2)$ into the eigenvector basis. Furthermore, because the last $n - k$ eigenvalues are zero, the same holds for the last $n - k$ entries of $\Lambda \mathbf{a}$ and $\Lambda \mathbf{b}$. This implies that the estimate of \mathbf{w} will be unbiased independently of the values of the last $n - k$ entries of \mathbf{a} and \mathbf{b} . It follows that the only requirement for an unbiased \mathbf{w} is that the projections of the mean estimates into the subspace of eigenvectors of non-zero eigenvalue be unbiased up to some common vector,

$$\begin{aligned} E_{X_1, \dots, X_{2n}}[\Psi \hat{\mu}_1] &= \Psi \mu_1 + \alpha \\ E_{X_1, \dots, X_{2n}}[\Psi \hat{\mu}_2] &= \Psi \mu_2 + \alpha \end{aligned}$$

where α is some non-zero vector and Ψ a $k \times d$ matrix whose rows are the eigenvectors of Σ^{-1} of non-zero eigenvalue.