Homework Set Four Solutions

ECE 175

Department of Electrical and Computer Engineering
University of California, San Diego

Nuno Vasconcelos

1. The log-likelihood is

$$L = \sum_i \left\{ \log \frac{n!}{x_i!(n-x_i)!} + x_i \log p + (n - x_i)\log(1-p) \right\}$$

and setting the derivative with respect to p to zero we get

$$\frac{1}{p}\sum_i x_i - \frac{1}{1-p}\sum_i (n - x_i) = 0$$

$$(1-p)\sum_i x_i = p \sum_i (n - x_i)$$

$$p = \frac{1}{N}\sum_i \frac{x_i}{n}$$

We next check the second-order derivative, which is

$$\frac{\partial^2 L}{\partial p^2} = -\frac{1}{p^2}\sum_i x_i - \frac{1}{(1-p)^2}\sum_i (n - x_i).$$

This is clearly not positive, since the $0 \leq x_i \leq n$ by definition of the binomial random variable (where x is the count of the number of successes in n Bernoulli trials of parameter p). Hence, we have a maximum.

2. a) Since $\mathbf{Ax}$ is a constant and $\mathbf{n}$ a Gaussian, Y is a Gaussian. The mean is $\mathbf{Ax} + \mathbf{0} = \mathbf{Ax}$ and the covariance $\mathbf{\Sigma}$.

b) The ML estimate is

$$\begin{aligned}
\hat{\mathbf{x}} &= \arg\min_{\mathbf{x}} (\mathbf{y} - \mathbf{Ax})^T \mathbf{\Sigma}^{-1} (\mathbf{y} - \mathbf{Ax}) + N \log(2\pi|\mathbf{\Sigma}|) \\
&= \arg\min_{\mathbf{x}} (\mathbf{y} - \mathbf{Ax})^T \mathbf{\Sigma}^{-1} (\mathbf{y} - \mathbf{Ax}) \\
&= \arg\min_{\mathbf{x}} (\mathbf{Sy} - \mathbf{SAx})^T (\mathbf{Sy} - \mathbf{SAx}) \\
&= \arg\min_{\mathbf{x}} ||\mathbf{Sy} - \mathbf{SAx}||^2
\end{aligned}$$

where $\mathbf{S}$ is the square root of $\mathbf{\Sigma}^{-1}$, i.e.

$$\mathbf{S}^T\mathbf{S} = \mathbf{\Sigma}^{-1}.$$

Note that this is the same as

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} ||\mathbf{v} - \mathbf{Mx}||^2$$

where $\mathbf{v} = \mathbf{Sy}$ and $\mathbf{M} = \mathbf{SA}$. The gradient with respect to $\mathbf{x}$ is zero when

$$\mathbf{M}^T(\mathbf{v} - \mathbf{Mx}) = 0$$

or

$$\hat{\mathbf{x}} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{v} = (\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}.$$

Since both $\mathbf{M}$ and $\mathbf{A}$ are positive definite, $\mathbf{M}$ is positive definite and we have a minimum, which is a maximum of the log-likelihood.

c) Note from **b)** that this is equivalent to the least squares problem

$$\hat{\mathbf{x}} \quad = \quad \arg\min_{\mathbf{x}} (\mathbf{y} - \mathbf{A}\mathbf{x})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x})$$

and when $\boldsymbol{\Sigma} = diag(\sigma_1^2, \dots, \sigma_n^2)$ this is the same as

$$\hat{\mathbf{x}} \quad = \quad \arg\min_{\mathbf{x}} \sum_i \frac{(y_i - (\mathbf{A}\mathbf{x})_i)^2}{\sigma_i^2}.$$

This is a weighted least squares cost, where the error in each component is weighed by the inverse of the variance of the noise. That is, we give more importance to the dimensions that have lower variance, i.e. where we are most sure that the data will be less noisy, and less weight to the dimensions where the noise dominates.

3. From the solution of problem 2 we see that the ML estimate of the uncorrupted test pattern is

$$\hat{\mathbf{x}} = (\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}.$$

For the computer problem $\mathbf{A} = a\mathbf{I}$ and $\boldsymbol{\Sigma} = v\mathbf{I}$

$$\Rightarrow \hat{\mathbf{x}} = (\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} = (av^{-1}a)^{-1}av^{-1}\mathbf{y} = a^{-1}\mathbf{y}.$$

and the ML estimate of the scale parameter is

$$\begin{aligned} \hat{a} &= \arg\min_a (\mathbf{y} - \mathbf{A}\mathbf{x})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}) \\ &= \arg\min_a (v^{-1}(\mathbf{y} - a\mathbf{I}\mathbf{x})^T(\mathbf{y} - a\mathbf{I}\mathbf{x})) \\ &= \arg\min_a \sum_i (y_i - ax_i)^2 \end{aligned}$$

The derivative w.r.t a is zero when

$$\sum_i y_i x_i - a\sum_i x_i^2 = 0.$$

$$\hat{a} = \frac{\sum_i y_i x_i}{\sum_i x_i^2}$$

and the second derivative is always positive, hence we have a minimum.

1. The ML estimate of scale parameter for the sample images is 0.67
2. Least Square distance: P(Error|Class = i) for all $i = 0 \dots 9$.
   Total error rate, P(error) = 0.11
3. Euclidean Distance Metric: P(Error|Class = i) for all $i = 0 \dots 9$. Total error rate, P(error) = 0.21

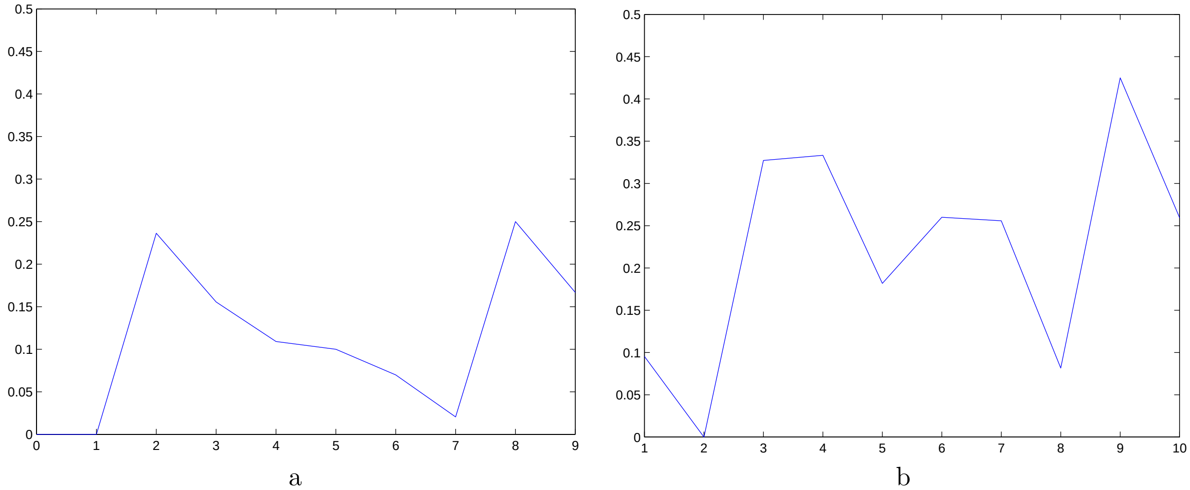

Figure 1: Probability of error for each class. a)Least Square metric b)Euclidean metric

Table 1: Probability of error for each class using Least Squares metric

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0.00	**0.00**	**0.23**	**0.16**	**0.11**	**0.10**	**0.07**	**0.02**	**0.25**	**0.17**

Table 2: Probability of error for each class using Euclidean metric

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0.10	**0.00**	**0.33**	**0.33**	**0.18**	**0.26**	**0.25**	**0.08**	**0.42**	**0.26**