Technion – Israel Institute of Technology *Summer 2014*

Electrical Engineering

Introduction to Machine Learning – International School Final Exam

1. Exam duration is **three hours**.
2. It is highly recommended to read the entire exam before you start.
3. Include brief explanations. You should answer all questions. The value of each question is given in the body of the question. The total number of points is 100.
4. You may use any material during the exam.
5. Write in a clear and organized manner.
6. The exam sheet includes 5 pages, including this page.

Good Luck

Distributions Table

Distribution	Notation	Support	PDF	Mean	Variance
Uniform	$x \sim U[a,b]$	$x \in [a,b]$	$f(x) = \frac{1}{b-a}$	$\frac{b+a}{2}$	$\frac{1}{12}(b-a)^2$
Normal	$x \sim N(\mu,\sigma^2)$	$\Re$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Exponential	$x \sim \text{Exp}(\lambda)$	$0 \le x \in \Re$	$f(x) = \frac{1}{\lambda} e^{-x/\lambda}$	λ	λ^2

Question 1 – Inference (29 points)

A decision problem is given, in which the input space is the real numbers, $X = \Re$. Input examples belong to one of three classes: $\Omega = \{e, g, u\}$. Class conditional distributions for the three classes are (see also distributions table in page 1 of the exam):

- Class u: $x \sim U[\lambda_u - 1, \lambda_u + 1]$
- Class g: $x \sim N(\lambda_g, 1)$
- Class e: $x \sim \mathrm{Exp}(\lambda_e)$

1. (7 points) Given are n independent samples $x_1, \ldots x_n$, drawn from class e. Write the maximum likelihood estimator (MLE) for the parameter λ_e of the exponential distribution. Is the estimator a biased estimator?

Assume until the remaining of the question that the parameters are known: $\lambda_u = \lambda_g = \lambda_e = 1$**.**

2. (7 points) Plot the conditional probability densities of the input given each one of the classes.
3. (8 points) Assuming the prior distribution over states is uniform, what is the optimal Bayes classifier of the state, given a single input x? Give a function from the real numbers to Ω.

Assume until the remaining of the question the prior distribution over states is $p(e) = 0.6$, $p(g) = p(u) = 0.2$**.**

4. (7 points) We are given n i.i.d observations $x_1 \ldots x_n$ from a **single** class, for a very large n. Define two random variables, the maximal value $x_{\max} = \max_i x_i$ and the minimal value $x_{\min} = \min_i x_i$. We look for a decision rule for predicting the state using the pair $[x_{\min}, x_{\max}]$. The rule should be of the form:

$$\begin{array}{ll} \text{if } (x_{\min} < \theta_A) & \text{then state is } \omega = \mathrm{A} \\ \text{else if } (x_{\max} > \theta_B) & \text{then state is } \omega = B \\ \text{else} & \text{state is } \omega = C \end{array}$$

Find a values for the thresholds $\theta_A, \theta_B \in \Re$ and the states $A, B, C \in \Omega$ such that the decision error given the state u is 0, and the decision error given the other 2 states is minimal.

Tutorial 14 : Question 1 of Summer 2014 Exam

1. Assuming $x_i \geq 0,\ i = 1, 2, ..., n$, the ML estimator is given by

$$\begin{aligned}\hat{\lambda}_{e\mathrm{MLE}} &= \arg\max_{\lambda} p(x_1, x_2, ..., x_n|\lambda) \\ &= \arg\max_{\lambda} \frac{1}{\lambda^n} \exp\left\{-\frac{1}{\lambda}\sum_{i=1}^{n} x_i\right\} \\ &= \arg\max_{\lambda}\ -n\log\lambda - \frac{1}{\lambda}\sum_{i=1}^{n} x_i \\ &= \arg\min_{\lambda} n\log\lambda + \frac{1}{\lambda}\sum_{i=1}^{n} x_i\end{aligned}$$

Setting the derivative w.r.t. λ to 0 we get

$$\frac{n}{\lambda} - \frac{1}{\lambda^2}\sum_{i=1}^{n} x_i = 0 \quad \Rightarrow \hat{\lambda}_{e\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Notice that $E[\hat{\lambda}_{e\mathrm{MLE}}] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \lambda$, hence, the estimator is unbiased.

Extra - The ML estimator of $\boldsymbol{\lambda_u}$:

The conditional probability density of a single sample x is given by

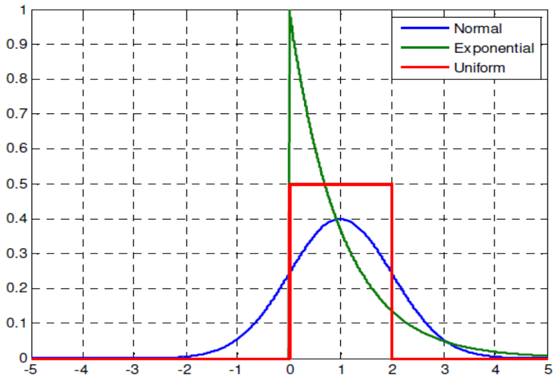
$$p(x|\lambda_u) = \begin{cases} \frac{1}{2}, & \lambda_u - 1 \leq x \leq \lambda_u + 1, \\ 0, & o.w. \end{cases}$$

Hence,

$$\begin{aligned}\hat{\lambda}_{u\mathrm{MLE}} &= \arg\max_{\lambda} p(x_1, x_2, ..., x_n|\lambda) \\ &= \arg\max_{\lambda} \begin{cases} \frac{1}{2^n}, & \lambda_u - 1 \leq x_i \leq \lambda_u + 1\ i = 1, 2, ..., n, \\ 0, & o.w. \end{cases} \\ &= \arg\max_{x_{\max}-1 \leq \lambda \leq x_{\min}+1} \frac{1}{2^n}\end{aligned}$$

In this case, the ML estimator is **not unique** and any number $\lambda \geq 0$ which satisfies the constraints can be considered as an ML estimator.

2. The conditional probabilities densities are as follows



3. Since the prior distributions over states is uniform (i.e. $p(u) = p(g) = p(e) = \frac{1}{3}$), we need to compare the conditional probabilities densities. Consider the following cases:

   - $x < 0$: In this case only the normal distribution has a positive density.
   - $0 \le x \le 2$: Here the normal distribution is smaller than the uniform distribution since $\frac{1}{\sqrt{2\pi}} \le \frac{1}{2}$ (see the plot), hence, we need to compare between the uniform and exponential distributions
   $$\exp\{-x\} \ge \frac{1}{2} \;\Rightarrow\; x \le \log 2.$$
   - $x > 2$: Now, the uniform distribution has zero density, hence, we compare between the normal and exponential distributions
   $$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(x-1)^2\} \le \exp\{-x\} \\ \Leftrightarrow\; & -\frac{1}{2}\log 2\pi - \frac{1}{2}(x-1)^2 \le -x \\ \Leftrightarrow\; & 0 \le \log 2\pi + x^2 - 2x + 1 - 2x = x^2 - 4x + (1 + \log 2\pi). \end{aligned}$$
   The roots are $x_{1,2} = 2 \pm \sqrt{3 - \log 2\pi}$ and only $x = 2 + \sqrt{3 - \log 2\pi}$ is relevant.

   Overall, the Bayes optimal classifier is given by the following function
   $$\hat{\omega}(x) = \begin{cases} g, & x < 0, \\ e, & 0 \le x \le \log 2, \\ u, & \log 2 \le x \le 2, \\ g, & 2 \le x \le 2 + \sqrt{3 - \log 2\pi}, \\ e, & x > 2 + \sqrt{3 - \log 2\pi}. \end{cases}$$

4. Now we need to compare the posterior probabilities. Consider the following cases:

   - $x < 0$: Again, in this case only the normal distribution has a positive density (thus, the prior distributions are irrelevant).
   - $0 \le x \le 2$: the normal and uniform distributions have the same prior probability and the uniform density is larger than the normal density, hence, we compare between the uniform and the exponential posterior probabilities
   $$0.6 \exp\{-x\} \ge 0.2 \cdot 0.5 \;\Rightarrow x \le \log 6.$$
   - $x > 2$: Again, the uniform density is 0, hence, we compare the normal and exponential posterior probabilities:
   $$\begin{aligned} &0.2 \cdot \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(x-1)^2\} \le 0.6 \cdot \exp\{-x\} \\ \Leftrightarrow\; & -\frac{1}{2}\log 2\pi - \frac{1}{2}(x-1)^2 \le log(3) - x \\ \Leftrightarrow\; & 0 \le \log 2\pi + x^2 - 2x + 12\log(3) - 2x = x^2 - 4x + (1 + \log 2\pi + 2\log 3). \end{aligned}$$
   The right-side term is always non-negative.

   Now, the Bayes optimal classifier is given by
   $$\hat{\omega}(x) = \begin{cases} g, & x < 0, \\ e, & 0 \le x \le \log 6, \\ u, & \log 6 \le x \le 2, \\ e, & x > 2. \end{cases}$$

   We require the error given the state u to be 0, hence, the decision for any sample in the range $[0, 2]$ must be u, which leads to the following rule:

   $$\begin{array}{ll} \text{if } (x_{\min} < 0) & \text{then state is } \omega = g \\ \text{else if } (x_{\max} > 2) & \text{then state is } \omega = e \\ \text{else} & \qquad\;\; \text{state is } \omega = u \end{array}$$

Extra

Compute the decision error given each one of the states and the average decision error. Is the decision rule is Bayes optimal for $n \to \infty$? Use that for $x \sim \mathcal{N}(1,1)$ we have $p(x > 0) \approx 0.84$.

The error given the state u is 0 of course. Given state g, we will have an error if $x_{\text{min}} > 0$ i.e., if all the samples are non-negative for which the probability is 0.84^n. For state e, we will have an error if all samples are in the range $[0, 2]$ for which the probability is $(1 - \exp\{-2\})^n$. Hence, the average decision error is

$$0.6 \cdot (1 - \exp\{-2\})^n + 0.2 \cdot 0.84^n \underset{n\to\infty}{\to} 0$$

This implies the decision rule is Bayes optimal.