Tsinghua-Berkeley Shenzhen Institute
INFORMATION THEORY AND STATISTICAL LEARNING
Fall 2020

Problem Set 3

Issued: Monday 19th October, 2020 **Due:** Friday 30th October, 2020

Notations: We use $\mathsf{x}, \mathsf{y}, \mathsf{w}$ and $\underline{\mathsf{x}}, \underline{\mathsf{y}}, \underline{\mathsf{w}}$ to denote random variables and random vectors. We use $\text{Bern}(p)$ to denote the Bernoulli distribution with the parameter p, and use $\text{Binom}(n, p)$ to denote the binomial distribution with parameters n and p.

3.1. (a) $P_{\mathsf{xy}}(x, y)$ is a joint distribution of discrete random variables x and y. Assume $x_0 \in \mathcal{X}$ is a value of x, prove that

$$I(\mathsf{x};\mathsf{y}) = \sum_{x\in\mathcal{X}} P_{\mathsf{x}}(x) D(P_{\mathsf{y}|\mathsf{x}=x}||P_{\mathsf{y}|\mathsf{x}=x_0}) - D(P_{\mathsf{y}}||P_{\mathsf{y}|\mathsf{x}=x_0})$$

(b) Let $\{P_{\mathsf{y}|\mathsf{x}=x}, x \in \mathcal{X}\}$ be a set of distributions. Prove that

$$\sup_{P_{\mathsf{x}}} I(\mathsf{x};\mathsf{y}) \leq \sup_{x,x'\in\mathcal{X}} D(P_{\mathsf{y}|\mathsf{x}=x}\|P_{\mathsf{y}|\mathsf{x}=x'}).$$

This is the information-theoretic version of "radius $\leq$ diameter".

Solution:

(a)

$$\begin{aligned}
\text{RHS} &= \sum_{x\in\mathcal{X}} P_{\mathsf{x}}(x) D(P_{\mathsf{y}|\mathsf{x}=x}||P_{\mathsf{y}|\mathsf{x}=x_0}) - D(P_{\mathsf{y}}||P_{\mathsf{y}|\mathsf{x}=x_0}) \\
&= \sum_{x\in\mathcal{X}} P_{\mathsf{x}}(x) \sum_{y\in\mathcal{Y}} P_{\mathsf{y}|\mathsf{x}=x}(y) \log\frac{P_{\mathsf{y}|\mathsf{x}=x}(y)}{P_{\mathsf{y}|\mathsf{x}=x_0}(y)} - \sum_{y\in\mathcal{Y}} P_{\mathsf{y}}(y)\log\frac{P_{\mathsf{y}}(y)}{P_{\mathsf{y}|\mathsf{x}=x_0}(y)} \\
&= \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} P_{\mathsf{xy}}(x,y)\log\frac{P_{\mathsf{y}|\mathsf{x}=x}(y)}{P_{\mathsf{y}|\mathsf{x}=x_0}(y)} - \sum_{y\in\mathcal{Y}} P_{\mathsf{y}}(y)\log\frac{P_{\mathsf{y}}(y)}{P_{\mathsf{y}|\mathsf{x}=x_0}(y)} \\
&= \sum_{y\in\mathcal{Y}} P_{\mathsf{y}}(y)\log\frac{1}{P_{\mathsf{y}}(y)} - \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} P_{\mathsf{xy}}(x,y)\log\frac{1}{P_{\mathsf{y}|\mathsf{x}=x}(y)} \\
&= H(\mathsf{y}) - H(\mathsf{y}|\mathsf{x}) \\
&= I(\mathsf{x};\mathsf{y}) \\
&= \text{LHS}
\end{aligned}$$

(b) Suppose when $P_{\mathsf{x}} = \tilde{P}_{\mathsf{x}}$, $I(\mathsf{x};\mathsf{y})$ derives the supremum.

$$\begin{aligned}
\sup_{P_{\mathsf{x}}} I(\mathsf{x};\mathsf{y}) &= \sum_{x\in\mathcal{X}} \tilde{P}_{\mathsf{x}}(x) D(P_{\mathsf{y}|\mathsf{x}=x}||P_{\mathsf{y}|\mathsf{x}=x_0}) - D(P_{\mathsf{y}}||P_{\mathsf{y}|\mathsf{x}=x_0}) \\
&\leq \sum_{x\in\mathcal{X}} \tilde{P}_{\mathsf{x}}(x) D(P_{\mathsf{y}|\mathsf{x}=x}||P_{\mathsf{y}|\mathsf{x}=x_0}) \\
&\leq \sup_{x'\in\mathcal{X}} \sum_{x\in\mathcal{X}} \tilde{P}_{\mathsf{x}}(x) D(P_{\mathsf{y}|\mathsf{x}=x}||P_{\mathsf{y}|\mathsf{x}=x'}) \\
&\leq \sup_{x,x'\in\mathcal{X}} D(P_{\mathsf{y}|\mathsf{x}=x}\|P_{\mathsf{y}|\mathsf{x}=x'})
\end{aligned}$$

3.2. (a) For discrete random variables x, y, z, prove

$$2H(\mathsf{x},\mathsf{y},\mathsf{z}) \leq H(\mathsf{x},\mathsf{y}) + H(\mathsf{y},\mathsf{z}) + H(\mathsf{z},\mathsf{x}).$$

(b) Use the above inequality to prove *Shearer's lemma*: Place n points in $\mathbb{R}^3$ arbitrarily. Let n_1, n_2, n_3 denote the number of distinct points projected onto the xy, xz and yz-plane, respectively. Then:

$$n_1 n_2 n_3 \geq n^2.$$

Solution:

(a) Think about the following facts:

$$H(\mathsf{x},\mathsf{y}) = H(\mathsf{x}) + H(\mathsf{y}|\mathsf{x})$$

$$H(\mathsf{y},\mathsf{z}) = H(\mathsf{y}) + H(\mathsf{z}|\mathsf{y})$$

$$H(\mathsf{z},\mathsf{x}) = H(\mathsf{z}) + H(\mathsf{x}|\mathsf{z})$$

$$H(\mathsf{x},\mathsf{y},\mathsf{z}) = H(\mathsf{x}) + H(\mathsf{y}|\mathsf{x}) + H(\mathsf{z}|\mathsf{x},\mathsf{y})$$

$$H(\mathsf{y}|\mathsf{x}) \leq H(\mathsf{y})$$

$$H(\mathsf{z}|\mathsf{x},\mathsf{y}) \leq H(\mathsf{z}|\mathsf{x})$$

$$H(\mathsf{z}|\mathsf{x},\mathsf{y}) \leq H(\mathsf{z}|\mathsf{y})$$

You can easily derive the inequality.

(b) $\{(x_i, y_i, z_i), i = 1, \cdots, n\}$ is a cardinality-n set. Each element has the same probability. Therefore,

$$H(\mathsf{x},\mathsf{y},\mathsf{z}) = \log n.$$

A similar description can be made on $H(\mathsf{x},\mathsf{y})$, $H(\mathsf{x},\mathsf{z})$, $H(\mathsf{y},\mathsf{z})$, but their elements are not definitely equiprobable. Therefore, $H(\mathsf{x},\mathsf{y}) \leq \log n_1$, $H(\mathsf{x},\mathsf{z}) \leq \log n_2$, $H(\mathsf{y},\mathsf{z}) \leq \log n_3$.

$$2\log n \leq \log n_1 + \log n_2 + \log n_3,$$

which leads to the *Shearer's lemma*.

3.3. Recall that $d(p||q) = D(\mathrm{Bern}(p)\|\mathrm{Bern}(q))$ denotes the binary divergence function:

$$d(p||q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \tag{1}$$

(a) Prove for all $p, q \in [0, 1]$

$$d(p||q) \geq 2(p-q)^2 \log e \tag{2}$$

(b) Apply data processing inequality (Chain Rule for K-L divergence) to prove the *Pinsker-Csiszatr inequality*:

$$\mathrm{TV}(P, Q) \leq \sqrt{\frac{1}{2\log e} D(P\|Q)} \tag{3}$$

where $\mathrm{TV}(P, Q)$ is the *total variation* distance between probability distribution P and Q:

$$\mathrm{TV}(P, Q) \triangleq \sup_{E\in\mathcal{F}}(P(E) - Q(E)), \tag{4}$$

with the supremum taken over all events E.

Solution:

(a) If p is taken as a constant,

$$f(q) = \mathrm{LHS} - \mathrm{RHS} = d(p||q) - 2(p-q)^2 \log e$$

Then,

$$f'(q) = (p-q)(4 - \frac{1}{q(1+q)}) \log e.$$

Since $4 \leq \frac{1}{q(1+q)}$,

$$f'(q) = \begin{cases} \geq 0 & q > p \\ \leq 0 & q < p \end{cases}.$$

Therefore,

$$f(q) \geq f(p) = 0,$$

which means LHS $\leq$ RHS.

(b) Let $E^+ = \{e|P(e) \geq Q(e)\}$,

$$P_{E^+} = \begin{cases} 1 & \text{w.p. } \sum\limits_{e\in E^+} P(e) \\ 0 & \text{w.p. } \sum\limits_{e\notin E^+} P(e) \end{cases}, \text{ and } Q_{E^+} = \begin{cases} 1 & \text{w.p. } \sum\limits_{e\in E^+} Q(e) \\ 0 & \text{w.p. } \sum\limits_{e\notin E^+} Q(e) \end{cases}.$$

It's easy to verify that $\mathrm{TV}(P, Q) = \mathrm{TV}(P_{E^+}, Q_{E^+})$.

Then, let $\mathsf{z} = \begin{cases} 1 & e \in E^+ \\ 0 & e \notin E^+ \end{cases}$. Since z is a function of e, we can also think of the two distributions P and Q as joint distributions for the random variables (e, z). By (a), applying the chain rule for KL-divergence gives

$$\begin{aligned} D(P_{\mathsf{ez}}\|Q_{\mathsf{ez}}) &= D(P_{\mathsf{z}}\|Q_{\mathsf{z}}) + D(P_{\mathsf{e|z}}\|Q_{\mathsf{e|z}}) \\ &\geq D(P_{\mathsf{z}}\|Q_{\mathsf{z}}) \\ &= D(P_{E^+}\|Q_{E^+}) \\ &\geq 2(\sum_{e\in E^+} P(e) - \sum_{e\in E^+} Q(e))^2 \log e \end{aligned}$$

That's the inequality.

3.4. Let y be a continuous random variable distributed over the closed interval $[0, 1]$. Under the null hypothesis H_0, y is uniform:

$$p_{\mathsf{y}|\mathsf{H}}(y|H_0) = \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{o.w.} \end{cases}$$

Under the alternative hypothesis H_1, the conditional pdf of y is as follows:

$$p_{\mathsf{y}|\mathsf{H}}(y|H_1) = \begin{cases} 2y, & 0 \le y \le 1 \\ 0, & \text{o.w.} \end{cases}$$

The *a-priori* probability that y is uniformly distributed is p.

(a) Find the decision rule that minimizes the expected error.

(b) Find the closed form expression for the operating characteristic of the LRT, i.e., $P_{\mathrm{D}} \triangleq \mathbb{P}(\hat{\mathsf{H}} = H_1 | \mathsf{H} = H_1)$ as a function of $P_{\mathrm{F}} \triangleq \mathbb{P}(\hat{\mathsf{H}} = H_1 | \mathsf{H} = H_0)$ for the likelihood ratio test.

(c) Suppose we require that P_{D} is at least $(1+\varepsilon)P_{\mathrm{F}}$, where $\epsilon > 0$ is a fixed constant.

i. Find $P_{\mathrm{D}}^{\max}(\varepsilon)$, the maximal value of P_{D} that is achievable under this constraint.

ii. Find the range of values of ε that lead to non-trivial performance, i.e. $P_{\mathrm{D}}^{\max}(\varepsilon) > 0$.

iii. When using the decision rule from part a, what values of p guarantee that $P_{\mathrm{D}} \ge (1+\varepsilon)P_{\mathrm{F}}$?

Solution:

(a) LRT becomes

$$\frac{p_{\mathsf{y}|H}(y|H_1)}{p_{\mathsf{y}|H}(y|H_0)} \underset{\hat{\mathsf{H}}(y)=H_0}{\overset{\hat{\mathsf{H}}(y)=H_1}{\gtrless}} \frac{p}{1-p}.$$

It leads to

$$y \underset{\hat{\mathsf{H}}(y)=H_0}{\overset{\hat{\mathsf{H}}(y)=H_1}{\gtrless}} \frac{p}{2(1-p)}.$$

It is interesting to note that for $p > 2/3$, $\hat{H}(y)$ is always assigned to H_0.

(b) Let y_0 be the threshold in LRT. $P_F = \int_{y_0}^{1} \mathrm{d}y$ and $P_D = \int_{y_0}^{1} 2y\mathrm{d}y$. Therefore, P_D as a function of P_F:

$$P_D(P_F) = (2 - P_F)P_F$$

(c) i. The maximal value of P_D that still satisfies the constraint is achieved at the point of intersection of the operating characteristic curve and the line $(1 + \epsilon)P_F$. Lets find this point. $P_D = (2 - P_F)P_F = (1 + \epsilon)P_F$. Substituting back to the equation for the operating characteristic curve, or the constraint, we get $P_D^{\max}(\epsilon) = 1 - \epsilon^2$.

ii. From part (c)(i), we conclude that $P_D^{max} > 0$ can only be obtained if the constraint line is below the operating characteristic curve at $P_F = 0$. Therefore, we need to find the conditions under which the slope of the constraint line is lower than the slope of the tangent to the operating characteristic curve at $P_F = 0$.
The equation of the tangent to $P_D = (2 - P_F)P_F$ is $P_D = 2 - 2P_F$. The slope of the tangent at $P_F = 0$ is therefore 2. Comparing this to the slope of the constraint line, $1 + \epsilon$, we obtain < 1. Finally, the range is $\epsilon \in [0, 1)$.

iii. We know from (c)(i) that under the constraint $P_D \geq (1 + \epsilon)P_F$, the minimum P_D we can obtain is zero and the maximum P_D we can obtain is $1 - \epsilon^2$. To find the p that results in the maximum P_D we set, $P_D = 1 - y_0^2 \leq 1 - \epsilon^2$. Also with the case $p > 2/3$, finally we get $p \in \left[\min\left\{\frac{2\epsilon}{1+2\epsilon}, \frac{2}{3}\right\}, 1\right]$.

3.5. A 3-dimensional random vector $\underline{\mathsf{y}}$ is observed, and we know that one of the three hypotheses is true:

$$\begin{aligned} H_1: &\quad \underline{\mathsf{y}} = \underline{m}_1 + \underline{\mathsf{w}} \\ H_2: &\quad \underline{\mathsf{y}} = \underline{m}_2 + \underline{\mathsf{w}} \\ H_3: &\quad \underline{\mathsf{y}} = \underline{m}_3 + \underline{\mathsf{w}}, \end{aligned}$$

where

$$\underline{\mathsf{y}} = \begin{bmatrix} \mathsf{y}_1 \\ \mathsf{y}_2 \\ \mathsf{y}_3 \end{bmatrix}, \quad \underline{m}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \underline{m}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \underline{m}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

and $\underline{\mathsf{w}}$ is a zero-mean Gaussian vector with covariance matrix $\sigma^2\mathbf{I}$.

(a) Let

$$\underline{\pi}(\underline{y}) = \begin{bmatrix} \mathbb{P}(\mathsf{H} = H_1 | \, \underline{\mathsf{y}} = \underline{y}) \\ \mathbb{P}(\mathsf{H} = H_2 | \, \underline{\mathsf{y}} = \underline{y}) \\ \mathbb{P}(\mathsf{H} = H_3 | \, \underline{\mathsf{y}} = \underline{y}) \end{bmatrix} = \begin{bmatrix} \pi_1(\underline{y}) \\ \pi_2(\underline{y}) \\ \pi_3(\underline{y}) \end{bmatrix},$$

and suppose that the Bayes costs are

$$C_{11} = C_{22} = C_{33} = 0, \quad C_{12} = C_{21} = 1, \quad C_{13} = C_{31} = C_{23} = C_{32} = 2.$$

i. Specify the optimum decision rule in terms of $\pi_1(\underline{y}), \pi_2(\underline{y})$ and $\pi_3(\underline{y})$.

ii. Recalling that $\pi_1 + \pi_2 + \pi_3 = 1$, express this rule completely in terms of π_1 and π_2, and sketch the decision regions in the (π_1, π_2) plane.

(b) Suppose that the three hypotheses are equally likely a priori and that the Bayes costs are

$$C_{ij} = 1 - \delta_{ij} = \begin{cases} 1, & i \neq j \\ 0, & i = j \end{cases}.$$

Show that the optimum decision rule can be specified in terms of the pair of sufficient statistics

$$\ell_2(\underline{\mathsf{y}}) = \mathsf{y}_2 - \mathsf{y}_1,$$
$$\ell_3(\underline{\mathsf{y}}) = \mathsf{y}_3 - \mathsf{y}_1.$$

Hint: To begin, see if you can specify the optimum decision rules in terms of

$$L_i(\underline{y}) = \frac{p_{\underline{\mathsf{y}}|\mathsf{H}}(\underline{y}|H_i)}{p_{\underline{\mathsf{y}}|\mathsf{H}}(\underline{y}|H_1)}, \quad \text{for } i = 2, 3.$$

Solution:

(a) i. The expected costs $\phi_1(\underline{y}), \phi_2(\underline{y}), \phi_3(\underline{y})$ of deciding H_1, H_2, H_3 are

$$\begin{bmatrix} \phi_1(\underline{y}) \\ \phi_2(\underline{y}) \\ \phi_3(\underline{y}) \end{bmatrix} \triangleq \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \begin{bmatrix} \pi_1(\underline{y}) \\ \pi_2(\underline{y}) \\ \pi_3(\underline{y}) \end{bmatrix} = \begin{bmatrix} \pi_2(\underline{y}) + 2\pi_3(\underline{y}) \\ \pi_1(\underline{y}) + 2\pi_3(\underline{y}) \\ 2\pi_1(\underline{y}) + 2\pi_2(\underline{y}) \end{bmatrix}.$$

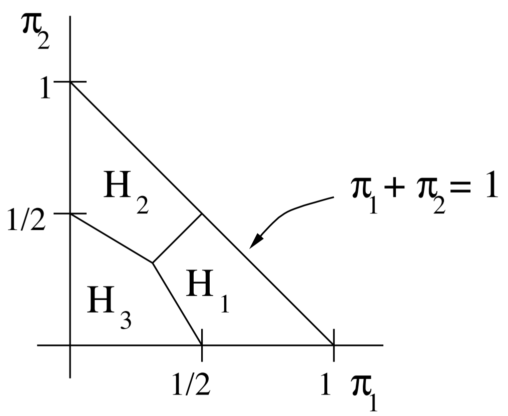
Then we have

$$\hat{\mathsf{H}}(\underline{y}) = H_j$$

with

$$j \triangleq \underset{i \in \{1,2,3\}}{\arg\min}\ \phi_i(\underline{y}).$$

ii.



(b) Since $C_{ij} = 1 - \delta_{ij}$ and the hypotheses are equally likely a priori, the ML rule is optimal. Therefore, we have

$$\hat{\mathsf{H}}(\underline{y}) = H_j$$

with

$$j \triangleq \underset{i \in \{1,2,3\}}{\arg\max}\ p_{\underline{\mathsf{y}}|\mathsf{H}}(\underline{y}|H_i),$$

which is equivalent to

$$\hat{\mathsf{H}}(\underline{y}) = \begin{cases} H_1, & L_2(\underline{y}) \leq 1 \text{ and } L_3(\underline{y}) \leq 1 \\ H_2, & L_2(\underline{y}) > 1 \text{ and } L_3(\underline{y}) \leq L_2(\underline{y}) \\ H_3, & L_3(\underline{y}) > 1 \text{ and } L_3(\underline{y}) > L_2(\underline{y}). \end{cases}$$

Then, since

$$L_i(\underline{y}) = \exp\left(\frac{\ell_i(\underline{y})}{\sigma^2}\right), \quad i = 2, 3,$$

the decision rule can be rewritten as

$$\hat{\mathsf{H}}(\underline{y}) = \begin{cases} H_1, & \ell_2(\underline{y}) \leq 0 \text{ and } \ell_3(\underline{y}) \leq 0 \\ H_2, & \ell_2(\underline{y}) > 0 \text{ and } \ell_3(\underline{y}) \leq \ell_2(\underline{y}) \\ H_3, & \ell_3(\underline{y}) > 0 \text{ and } \ell_3(\underline{y}) > \ell_2(\underline{y}). \end{cases}$$

3.6. A binary random variable x with prior $p_{\mathsf{x}}(\cdot)$ takes values in $\{-1, 1\}$. It is observed via n separate sensors; y_i denotes the observation at sensor i. The $\mathsf{y}_1, \cdots, \mathsf{y}_n$ are conditionally independent given x, i.e.,

$$p_{\mathsf{y}_1,\cdots,\mathsf{y}_n|\mathsf{x}}(y_1, \cdots, y_n|x) = \prod_{i=1}^{n} p_{\mathsf{y}_i|\mathsf{x}}(y_i|x).$$

A *local* decision $\hat{x}_i(y_i) \in \{-1, 1\}$ about the value of x is made at each sensor.

(a) In this part of the problem, each sensor sends its local decision to a fusion center. The fusion center combines the local decisions from all sensors to produce a global decision $\hat{x}(\hat{x}_1, \cdots, \hat{x}_n)$. Consider the special case in which:

- $P_{\mathsf{x}}(1) = P_{\mathsf{x}}(-1) = 1/2$;
- $\mathsf{y}_i = \mathsf{x} + \mathsf{w}_i$, where $\mathsf{w}_1, \cdots, \mathsf{w}_n$ are independent and each uniformly distributed over the interval $[-2, 2]$;
- the local decision rule is a simple thresholding of the observation, i.e.,

$$y_i \underset{\hat{x}_i(y_i)=-1}{\overset{\hat{x}_i(y_i)=1}{\gtrless}} 0.$$

Determine the minimum probability of error decision $\hat{x}(\cdot, \ldots, \cdot)$, at the fusion center.

In the remainder of the problem, there is no fusion center. The prior $P_{\mathsf{x}}(\cdot)$, observation model $p_{\mathsf{y}_i|\mathsf{x}}(\cdot|x), i = 1, 2$, and local decision rules $\hat{x}_i$, are no longer restricted as in part (a). However, we limit our attention to the two-sensor case ($n = 2$).

Consider local decisions $\hat{x}_i(y_i), i = 1, 2$, that minimize the expected cost, where the cost is defined for the two local rules jointly. Specifically, $C(\hat{x}_1, \hat{x}_2, x)$ is the cost of deciding $\hat{x}_1$ at sensor 1 and deciding $\hat{x}_2$ at sensor 2 when the true value of x is x. The cost C strictly increases with the number of errors made by the two sensors but is not necessarily symmetric.

(b) First, assume $\hat{x}_2(\cdot)$ is given. Show that the choice $\hat{x}_1^*(\cdot)$ for $\hat{x}_1(\cdot)$ that minimizes the expected (joint) cost is a likelihood ratio test of the form

$$\frac{p_{\mathsf{y}_1|\mathsf{x}}(y_1|1)}{p_{\mathsf{y}_1|\mathsf{x}}(y_1|-1)} \underset{\hat{x}_1(y_1)=-1}{\overset{\hat{x}_1(y_1)=1}{\gtrless}} \gamma_1.$$

where γ_1 is a threshold that depends on the rule $\hat{x}_2(\cdot)$. Determine the threshold γ_1.

(c) Assuming, instead, that $\hat{x}_1(\cdot)$ is given, determine the choice $\hat{x}_2^*(\cdot)$ for $\hat{x}_2(\cdot)$ that minimizes the expected joint cost.

(d) Consider a joint cost function $C(\hat{x}_1, \hat{x}_2, x)$ such that the cost is: 0 if both sensors making correct decisions; 1 if exactly one sensor makes a mistake; and L if both sensors make an error. Determine the value of L such that the optimal local decision rules at the two sensors are decoupled, i.e., the optimal threshold γ_1 does not depend on $\hat{x}_2^*(\cdot)$, and *vice versa*.

Solution:

(a) Since $\mathsf{w}_1, \ldots, \mathsf{w}_n$ are independent and uniform over the interval $[-2, 2]$ we have

$$p_{\hat{\mathsf{x}}_i|\mathsf{x}}(1|1) = p_{\hat{\mathsf{x}}_i|\mathsf{x}}(-1|-1) = \frac{3}{4}$$

$$p_{\hat{\mathsf{x}}_i|\mathsf{x}}(-1|1) = p_{\hat{\mathsf{x}}_i|\mathsf{x}}(1|-1) = \frac{1}{4}$$

Denoting $n_1 = \sum_i \frac{1}{2}(\hat{x}_i + 1)$, i.e., the number of sensors with a local decision of $\hat{x}_i = 1$, we have the ML decision rule

$$\frac{\left(\frac{3}{4}\right)^{n_1}\left(\frac{1}{4}\right)^{n-n_1}}{\left(\frac{3}{4}\right)^{n-n_1}\left(\frac{1}{4}\right)^{n_1}} \underset{\hat{x}=-1}{\overset{\hat{x}=1}{\gtrless}} 1$$

Finally it will give

$$\sum_{i=1}^{n} \hat{x}_i \underset{\hat{x}=-1}{\overset{\hat{x}=1}{\gtrless}} 0.$$

(b) $\gamma_1 = \dfrac{P_{\mathsf{x}}(-1)\,\mathbb{E}\left[C(1, \hat{x}_2(\mathsf{y}_2), -1) - C(-1, \hat{x}_2(\mathsf{y}_2), -1)|\mathsf{x} = -1\right]}{P_{\mathsf{x}}(1)\,\mathbb{E}\left[C(-1, \hat{x}_2(\mathsf{y}_2), 1) - C(1, \hat{x}_2(\mathsf{y}_2), 1)|\mathsf{x} = 1\right]}$

(c) $\dfrac{p_{\mathsf{y}_2|\mathsf{x}}(y_2|1)}{p_{\mathsf{y}_2|\mathsf{x}}(y_2|-1)} \underset{\hat{x}_2(y_2)=-1}{\overset{\hat{x}_2(y_2)=1}{\gtrless}} \dfrac{P_{\mathsf{x}}(-1)\,\mathbb{E}\left[C(\hat{x}_1(\mathsf{y}_1), 1, -1) - C(\hat{x}_1(\mathsf{y}_1), -1, -1)|\mathsf{x} = -1\right]}{P_{\mathsf{x}}(1)\,\mathbb{E}\left[C(\hat{x}_1(\mathsf{y}_1), -1, 1) - C(\hat{x}_1(\mathsf{y}_1), 1, 1)|\mathsf{x} = 1\right]}.$

(d) Compute γ_1. Since $p_{\hat{\mathsf{x}}_2|\mathsf{x}}(\hat{x}_2|x)$ depends on the second sensors decision rule, if we want the threshold to be independent of this rule for any likelihood model, we have to pick $L = 2$.