These are exercises for self-study and exam preparation. All material is examinable unless otherwise mentioned.

Exercise 1. ***Optimal actions for different loss functions***

Assume that $h \in \{1, \ldots, 5\}$ with probabilities $f(h)$ equal to

$$f(h=1)=0.2, \quad f(h=2)=0.3, \quad f(h=3)=0, \quad f(h=4)=0.1, \quad f(h=5)=0.4 \tag{1}$$

Compute the action a that minimises the expected loss $\mathbb{E}_{f(h)}[\ell(h,a)]$ when:

(a) $\ell(h,a)=(h-a)^2$

(b) $\ell(h,a)=|h-a|$

(c) $\ell(h,a)=1-\mathbb{1}(h=a)$

(d) $\ell(h,a)=-\log a(h)$ where choosing an action a means choosing a distribution over h.

Exercise 2. ***Causal decision theory applied to the kidney stone example***

The table summarizes outcomes for two types of surgery ($T=a$ and $T=b$) to remove kidney stones. Success rates are presented both overall and by stone size. For each treatment and size category, the table also reports the number of successes ($R=1$) and the number of patients treated. For example, 192/263 means that 263 patients received treatment a and for 192 of them the surgery was successful ($R=1$).

	Overall success rate	Small stones	Large stones
Treatment a	78% (273/350)	93% (81/87)	73% (192/263)
Treatment b	83% (289/350)	87% (234/270)	69% (55/80)

To decide which treatment to choose when the size of the kidney stone is unknown, we specify the following loss function:

$$\ell_0(R) = \begin{cases} c & \text{if } R=0 \\ 0 & \text{if } R=1 \end{cases} \tag{2}$$

where $c>0$. The loss ignores recovery time and other considerations, and simply assigns a penalty $c>0$ to an unsuccessful surgery.

(a) Determine the treatment T that minimises

$$\mathcal{R}(T) = \mathbb{E}_{p(R;do(T))}\left[\ell_o(R)\right]. \tag{3}$$

You may use that

$$p(R=1; do(T)=a) = 0.833 \qquad p(R=1; do(T)=b) = 0.779 \tag{4}$$

(b) Assume now that the success rates are considered uncertain (e.g because they are computed from a small number of patients only). How can this uncertainty be incorporated in the decision making?

Exercise 3. ***Classification with an asymmetric loss***

We here derive the optimal policy for binary classification when the costs of false positives and false negatives are unequal. Denote by $h \in \{0, 1\}$ the true class label and the predicted label by $a \in \{0, 1\}$. We consider the following loss

h	a	$\ell(h, a)$
0	0	0
1	0	c_{fn}
0	1	c_{fp}
1	1	0

where $c_{fn} > 0$ indicates the cost of a false negative and $c_{fp} > 0$ the cost of a false positive. We assume that we were given data $\mathbf{x}$ and that can compute the posterior $p(h|\mathbf{x})$. Derive the policy $a^*(\mathbf{x})$ that minimises the posterior expected loss, i.e.

$$a^*(\mathbf{x}) = \underset{a}{\operatorname{argmin}} \, \mathbb{E}_{p(h|\mathbf{x})} \left[\ell(h, a) \right] \tag{5}$$

Exercise 4. ***Penalised squared loss***

We consider the squared loss subject to a squared penalty for the action to deviate from zero,

$$\ell(h, a) = (h - a)^2 + \lambda a^2, \tag{6}$$

where $\lambda > 0$ is a weighting factor.

Derive the action that minimise the expected loss $\mathbb{E}_{f(h)} \left[\ell(h, a) \right]$.