SOLUTION OF HOMEWORK
Decision Theory
(MacKay - Chapter 36)

Necessary reading for this assignment:

- *Information Theory, Inference, and Learning Algorithms* (MacKay):
  - Chapter 36.1: *Rational prospecting*
  - Chapter 36.2: *Further reading*
- *Additional material:*
  - *Slide-set of the lecture on Decision Theory*

Note: The exercises are labeled according to their level of difficulty: `[Easy]`, `[Medium]` or `[Hard]`. This labeling, however, is subjective: different people may disagree on the perceived level of difficulty of any given exercise. Don't be discouraged when facing a hard exercise, you may find a solution that is simpler than the one the instructor had in mind!

Review questions.

1. Answer formally the following questions. In a decision theory problem:

   (a) What do the set of *states of the world* model?

   Instructor's solution: The set of states of the world models all states in which reality can be.

   (b) What do the set of *actions* model?

   Instructor's solution: The set of actions model the actions a person (or entity) can take.

   (c) What does the *utility function* model?

   Instructor's solution: The utility function models how beneficial it is for the person (or entity) to take each certain action if the world is in each state. The higher the value of the function for a pair (a, x), the more beneficial it is to take action a when the state of the world is x.

   (d) What is the quantity we are trying to maximize?

   Instructor's solution: We are trying to maximize the expectation

   $$E[U \mid a] = \sum_{a} p(x \mid a) u(a, x),$$

   of the utility of an action, where a ranges over actions, x ranges over states of the world, $p(x \mid a)$ is the probability that the world is in state x given that we took action a, and $u(a, x)$ is the utility of taking action a if the world is in state x.
   We are free to pick an action a that maximizes the above quantity.

Exercises.

2. `[Medium]` (The bicycle shop.) Zed and Adrian and run a small bicycle shop called "Z to A Bicycles". They must order bicycles for the coming season. Orders for the bicycles must be placed in quantities multiple of twenty (20). The cost per bicycle is \$70 if they order 20, \$67 if they order 40, \$65 if they order 60, and \$64 if they order 80. The bicycles will be sold for \$100 each. Any bicycles left over at the end of the season can be sold (for certain) at \$45 each. If Zed and Adrian run out of bicycles during the season, then they will suffer a loss of "goodwill" among their customers. They estimate this goodwill loss to be \$5 per customer who was unable to buy a bicycle. Zed and Adrian estimate that the demand for bicycles this season will be 10, 30, 50, or 70 bicycles with probabilities of 0.2, 0.4, 0.3, and 0.1 respectively.

   Decide Zed and Adrian's best possible action in this scenario, that is, determine the amount of bicycles they must order for the coming season in order to maximize their expected profit.

   Instructor's solution: There are four actions available to Zed and Adrian:

   (i) buy 20 bicycles,
   (ii) buy 40 bicycles,
   (iii) buy 60 bicycles, and
   (iv) buy 80 bicycles.

   There are four possible states of nature:

   (i) the demand is 10 bicycles,
   (ii) the demand is 30 bicycles,
   (iii) the demand is 50 bicycles, and
   (iv) the demand is 70 bicycles.

   Zed and Adrian have control over which action they choose, but they have no control over which state of nature will occur. They can only plan and make the best decision.

   To compute the utility function U, we need to check what is the payoff to Zed and Adrian for each possible combination of action taken and state of the world.

   For instance:

   - Utility for action "buy 60" when the state of the world is "demand is 50". They bought 60 at \$65 each for \$3 900. That is -\$3 900 since that is money they spent. Now, they sell 50 bicycles at \$100 each for \$5 000. They had 10 bicycles left over at the end of the season, and they sold those at \$45 each of \$450. That makes \$5 000 + \$450 - \$3 900 = \$1 550.
   - Utility for action "buy 40" when the state of the world is "demand is 70". They bought 40 at \$67 each for \$2 680. That is a negative \$2 680 since that is money they spent. Now, they sell 40 bicycles (that's all they had) at \$100 each for \$4 000. The other 30 customers that wanted a bicycle, but couldn't get one, left mad and Zed and Adrian lost \$5 in goodwill for each of them. That's 30 customers at -\$5 each or -\$150. That makes \$4 000 - \$2 680 - \$150 = \$1 170.

   The utility function U for every pair of action and state of the world is given in Table 1.

   Now, let us compute the expected value for each action and choose the action with the greatest expectation of utility.

   That is, let us pick the action that maximizes

$$E[U \mid a] = \sum_{x \in \mathcal{X}} p(x \mid a) U(x, a).$$

		States of the World		
	demand is 10	demand is 30	demand is 50	demand is 70
Actions buy 10	\$50	\$550	\$450	\$350
buy 40	-\$330	\$770	\$1 270	\$1 170
buy 60	-\$650	\$450	\$1 550	\$2 050
buy 80	-\$970	\$130	\$1 230	\$2 330

Tabela 1: Utilility function for the bicycle shop.

For that we need the probabilities of the states of the world given any action. But note that, in this problem, the state of the world is independent of the action taken by Zed and Aiden, and we have that for all actions a:

$$\begin{aligned} p(\text{demand is } 10 \mid a) &= 0.2, \\ p(\text{demand is } 30 \mid a) &= 0.4, \\ p(\text{demand is } 50 \mid a) &= 0.3, \qquad\qquad \text{and} \\ p(\text{demand is } 70 \mid a) &= 0.1. \end{aligned}$$

Now we compute the expected value for each action.

For instance, for action "buy 60" we get:

$$\begin{aligned} E[U \mid \text{buy } 60] = & \, p(\text{demand is } 10 \mid \text{buy } 60) \cdot U(\text{demand is } 10, \text{buy } 60) + \\ & p(\text{demand is } 30 \mid \text{buy } 60) \cdot U(\text{demand is } 30, \text{buy } 60) + \\ & p(\text{demand is } 50 \mid \text{buy } 60) \cdot U(\text{demand is } 50, \text{buy } 60) + \\ & p(\text{demand is } 70 \mid \text{buy } 60) \cdot U(\text{demand is } 70, \text{buy } 60) \\ = & \, 0.20 \cdot (-\$650) + 0.40 \cdot \$450 + 0.30 \cdot \$1\,150 + 0.10 \cdot \$2\,050 \\ = & \, \$720. \end{aligned}$$

Doing the same for the other actions we obtain:

$$\begin{aligned} E[U \mid \text{buy } 20] &= \$400, \\ E[U \mid \text{buy } 40] &= \$740, \qquad\qquad \text{and} \\ E[U \mid \text{buy } 80] &= \$460. \end{aligned}$$

Therefore we conclude that the best possible payoff is \$740, happening when Zed and Aiden take action "buy 40", and that's the decision they should make.

3. (MacKay 36.5) `[Hard]`

   (Hint: Model the problem as a decision theory problem (i.e., identify the states of the world, the possible actions, the probabilities involved) in which the utility function $U(x)$ is free. Then impose the restrictions that action A is preferred over B and that action D is preferred over C to derive a contradiction on the existence of $U(x)$.)

   Instructor's solution: Let us first carefully model the problem as decision problem.

   The possible states of the world are given by the set $\mathcal{X} = \{0, 1, 2.5\}$, representing how many million pounds can be granted as a prize.

The actions available are given by the set $\mathcal{A} = \{A, B, C, D\}$, representing the choices available for the person (we will call them "player") playing the "game".

In a decision theory framework, rational players use a utility function $u : \mathcal{A} \times \mathcal{X} \to \mathbb{R}$ indicating how much utility they obtain if they pick action $a \in \mathcal{A}$ and the world is in state $x \in \mathcal{X}$. In this particular exercise, we are told that the utility does not depend on the action taken, but only on the state of the world, so the utility function used is $U(x) : \mathcal{X} \to \mathbb{R}$.

The expected utility for the player given that they picked action $a\mathcal{A}$ is given by

$$E[U \mid a] = \sum_{x \in \mathcal{X}} p(x \mid a) U(x).$$

To calculate these quantities for each action, we need the probabilities summarized in Table 2.

States of the World

Actions	0	1	2.4
A	0	1	0
B	0.01	0.89	0.10
C	0.89	0.11	0
D	0.90	0.0	0.10

Tabela 2: Probabilities $p(x \mid a)$ of state x of the world given action a.

We can, then, calculate:

$$\begin{aligned} E[U \mid A] = &\sum_{x \in \mathcal{X}} p(x \mid A) U(x) \\ = &\, p(0 \mid A)U(0) + p(1 \mid A)U(1) + p(2.5 \mid A)U(2.5) \\ = &\, 0 \cdot U(0) + 1 \cdot U(1) + 0 \cdot U(2.5) \\ = &\, U(1) \end{aligned}$$

$$\begin{aligned} E[U \mid B] = &\sum_{x \in \mathcal{X}} p(x \mid B) U(x) \\ = &\, p(0 \mid B)U(0) + p(1 \mid B)U(1) + p(2.5 \mid B)U(2.5) \\ = &\, 0.01 \cdot U(0) + 0.89 \cdot U(1) + 0.10 \cdot U(2.5) \end{aligned}$$

$$\begin{aligned} E[U \mid C] = &\sum_{x \in \mathcal{X}} p(x \mid C) U(x) \\ = &\, p(0 \mid C)U(0) + p(1 \mid C)U(1) + p(2.5 \mid C)U(2.5) \\ = &\, 0.89 \cdot U(0) + 0.11 \cdot U(1) + 0 \cdot U(2.5) \\ = &\, 0.89 \cdot U(0) + 0.11 \cdot U(1) \end{aligned}$$

$$\begin{aligned} E[U \mid D] = &\sum_{x \in \mathcal{X}} p(x \mid D) U(x) \\ = &\, p(0 \mid D)U(0) + p(1 \mid D)U(1) + p(2.5 \mid D)U(2.5) \\ = &\, 0.90 \cdot U(0) + 0 \cdot U(1) + 0.10 \cdot U(2.5) \\ = &\, 0.90 \cdot U(0) + 0.10 \cdot U(2.5) \end{aligned}$$

We are told that players usually pick action A over B, and pick action D over C. That is, we are told that

$$E[U \mid A] > E[U \mid B] \qquad \text{and} \qquad E[U \mid D] > E[U \mid C],$$

and we have to show that there is no utility function $U(x)$ that would be consistent with this kind of player's choices.

According to our calculations:

$$\begin{aligned} E[U \mid A] > E[U \mid B] &\Rightarrow U(1) > 0.01 \cdot U(0) + 0.89 \cdot U(1) + 0.10 \cdot U(2.5) \\ &\Rightarrow U(0) - 11 \cdot U(1) + 10 \cdot U(2.5) < 0, \qquad (*) \end{aligned}$$

and

$$\begin{aligned} E[U \mid D] > E[U \mid C] &\Rightarrow 0.90 \cdot U(0) + 0.10 \cdot U(2.5) > 0.89 \cdot U(0) + 0.11 \cdot U(1) \\ &\Rightarrow U(0) - 11 \cdot U(1) + 10 \cdot U(2.5) > 0. \qquad (**) \end{aligned}$$

However, it is clear that $(*)$ and $(**)$ cannot be both true at the same time, no matter what $U(x)$ is, and therefore there is no utility function $U(x)$ that is consistent with the type of choice the player makes.

This is a sign that the player is not really rational in their decision making.