Bayesian decision with discrete probabilities

What we know:

A student visits open days at three different Universities. The student know that that the uni have done the following decisions about making applications to their degrees for students attending their open days.

- University 1 30% → 0,3
- University 2 50% → 0,5
- University 3 20% → 0,2

Further knows the student the likelyhood based on the analysis of the student experiences (X) in the open day if its Exciting or boring

- Uni 1 Exciting 0,7
- Uni 2 " 0,6
- Uni 3 " 0,3

What we seek:

The student would like to know where he shall apply for a degree.

1. First he wants to know what the probability of a exciting open day is, for all Universities.
2. Help to decide based on posterior probabilities, also show him the risk
3. The student asks himself how he could minimize the error (advise)
4. The student likes to know what the loss function is, can you show it to him?
5. Use the loss function to minimize the error rate and advise where he should apply.

likelyhood #3 ⟶ Prior probability #1

1) Formula

$$P(wi \mid x) = \frac{p(x \mid wi) \cdot p(wi)}{P(x)}$$

#4 Posterior Probability ⟵ Marginal likelihood #2 or evidence

#order to calculate

2) Fill in what you know example University 1

$$P(wi \mid x) = \frac{0,7 \times 0,3}{P(x)}$$

p(x) = Marginal likelyhood is missing

3) Formula (we know from week 2 - 18.19) Answer 1

Where in case of two categories the "evidence"

$$P(x) = \sum_{j=1}^{j=3} p(x \mid wj)\, P(wj) = P(\text{exciting}) = \sum_{\text{university } 1}^{\text{university } 3} p(\text{exciting}) \cdot p(\text{acceptance})$$

$$p(\text{exciting}) = \begin{bmatrix} 0,7 \times 0,3 \\ 0,6 \times 0,5 \\ 0,3 \times 0,2 \end{bmatrix} = 0,57$$

p(boring) only when two classes $1 - 0,57 = 0,43$

4) Fill into the formula (Uni 1) Answer 2

$$- \ p(\underset{\uparrow \text{ university}}{wi} \mid \overset{\text{the two classes} \downarrow}{\text{exciting}}) = \frac{0,7 \times 0,3}{0,57}$$

$$- \ p(U1 \mid \text{exciting}) = \frac{0,7 \times 0,3}{0,57} = 0,37$$

$$- \ p(U2 \mid \text{exciting}) = \frac{0,6 \times 0,5}{0,57} = 0,53$$

$$- \ p(U3 \mid \text{exciting}) = \frac{0,3 \times 0,2}{0,57} = 0,11$$

because there are two classes exciting & boring = 1

5) Help to decide

- We know that Decision given the posterior probabilities

  if $P(w_1|x) > P(w_2|x) \Rightarrow$ True state of nature $= w_1$

- Therefore we can order:

$$P(U_2|x) > P(U_1|x) > P(U_3|x)$$

- The posterior probability of University is the highest, thus we apply there.

6) Wrong choice ?!

• Wherever we observe a particular x, the probability of error is:

$$P(error|x) = P(w_1|x) \text{ if we decide } w_2$$
$$P(error|x) = P(w_2|x) \text{ if we decide } w_1$$

• Therefore

$$P(error|x) = \min[P(w_1|x), P(w_2|x)]$$

• Transform

$$P(error|x) = \min[P(U_1|x), P(U_2|x), P(U_3|x)]$$
$$\underline{\underline{0,10}} = \min[(0,37), (0,53), (0,10)]$$

The probability of making the wrong choice is at least 10%

(3) In case we want to minimize this error further, we can increase the numbers of samples, increase the dimension of feature space by using more features, thereby changing the likelihood $p(x|w_1)$ or combination of both

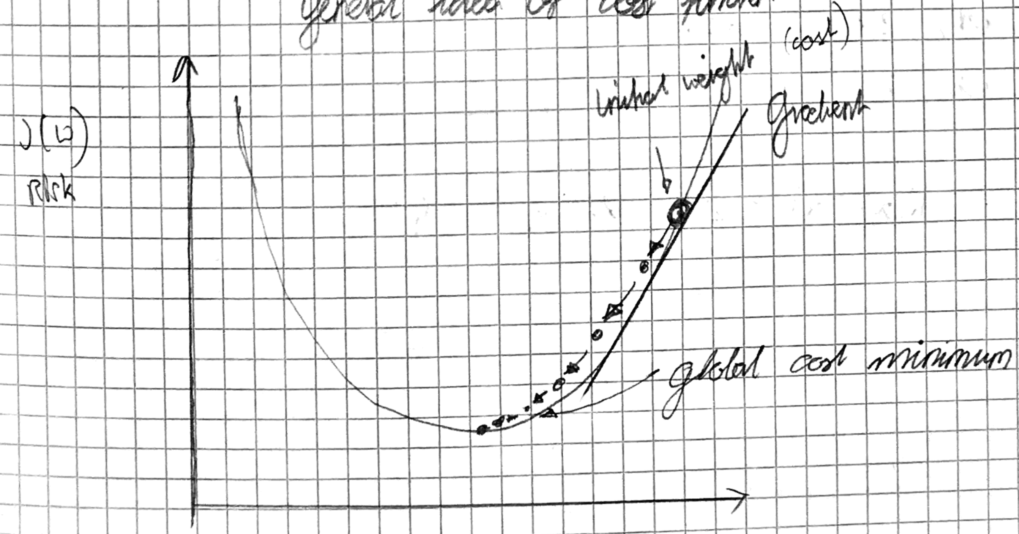
- Use more than one feature
- Use more than two states of nature
- Allowing actions that only decide on the state of nature
- Introduce a loss function which is more general than the probability of error.

(4) Introducing the loss function (cost function)

- allowing actions other than classification primarily allows the possibility or rejection
- Refusing to make a decision in close or bad cases!
- The loss function states how costly each action taken is.

The loss or cost function is usually used to find the global cost minimum. This is an optimization problem and in this context are called Bayes estimator. We can feed the model with more important features such as (costs) in this example and calculate the (local) (global) minimum

General idea of cost function



⑤ Theory

- let $\{w_1, w_2, w_3\}$ be the set of c states of nature "categories" Universities
- let $\{\alpha_1, \alpha_2, a_a\}$ be the set of possible actions
- let $\lambda(\alpha_i | w_j)$ be the loss incurred for taking action α_i when the state of nature is w_j.
- R = sum of all $R(\alpha_i|x)$ for i

1) You read on glassdor :

$U1 = \bar{\bar{5}} - 4K < \bar{\bar{5}}(U2) > \bar{\bar{5}}(U1)$

U1 = −4 k than U2

U2 = +1 k than U3

2) in 20 years that would mean

	U1	U2	U3
Cost	80	0	20

3) state of actions (apply α_1 = Reject α_2) α

4) final table (from teacher)

	U1	U2	U3
apply	80	0	100
rejection	60	90	0

5) Formula $R(\alpha_i | x) = \sum_{j=1}^{j=3} \lambda(\alpha_1 | w_j) P(w_j | x)$

$$R(\alpha_1 | x) = \begin{pmatrix} 80 \times 0{,}37 \\ 0 \times 0{,}57 \\ 100 \times 0{,}1 \end{pmatrix} = 39{,}6$$

$$R(\alpha_2 | x) = \begin{pmatrix} 60 \times 0{,}37 \\ 90 \times 0{,}53 \\ 0 \times 0{,}37 \end{pmatrix} = 69{,}9$$

6) Overall Risk
Select the action α for which $R(\alpha \mid x)$ is minimum in this case, R is called the 'Bayes' Risk which is the best performance that can be achieved.

Answer: Apply to a place in the university because over 20 years it will cost you more than to reject a university offer.

7) Improve decisions (self made)

	U1	U2	U3
Apply to U1	0	20	40
Apply to U2	80	0	20
Apply to U3	30	80	0

$R(\alpha_1 \mid x) = 0 \times 0{,}37 + 20 \times 0{,}53 + 40 \times 0{,}1 = 14{,}6$

$R(\alpha_2 \mid x) = 80 \times 0{,}37 + 0 \times 0{,}53 + 20 \times 0{,}1 = 31{,}6$

$R(\alpha_3 \mid x) = 30 \times 0{,}37 + 80 \times 0{,}53 + 0 \times 0{,}1 = 53{,}5$

8) According to this, which may sound counter intuitive you should apply to U1 and not to U2.