Solutions to Test 1 for Uncertainty in AI class, Fall 2025

1. Why cannot we just use machine learning to estimate the degree of confidence in its predictions – and to estimate how accurate are these predictions?

Answer. When we train a system to predict the result, we can (and do) use thousands of examples. However, when we train it to get the probability of each result, we need about 100 examples of computations to get one probability estimates – so the number of inputs for such training is 100 times smaller, not enough to get good results.

2a. If we know that the subjective probability of an event is between 0.6 and 0.8, and we want to find this probability with higher accuracy, what shall we do?

Answer. We find the midpoint of the interval – which is

$$\frac{0.6 + 0.8}{2} = 0.7,$$

and we ask the user which of the two options he/she prefers:

- to get \$100 if the event happens, and
- to get \$100 with probability 0.7.

If the user prefers the first option, this means that his/her subjective probability p of this event is higher than 0.7. So in this case, we can conclude that p is located in the half-width interval $[0.7, 0.8]$.

On the other hand, if the user prefers the second option, this means that his/her subjective probability p of this event is smaller than 0.7. So in this case, we can conclude that p is located in the half-width interval $[0.6, 0.7]$.

2b. How many questions do we need to ask a user to get his/her subjective probability with accuracy 5%?

Answer. Before we get any answers, all we know is that the probability is somewhere on the interval $[0, 1]$. The width of this interval is 1. After each answer, the width decreases by half. So, after k questions, the width becomes 2^{-k}. The smallest k for which 2^{-k} is smaller than 0.05 is $k = 5$ – then $2^{-k} = 1/32 < 1/20$. So, we need to ask 5 questions.

3. If 9 experts out of 12 think that the sock market will go down by the end of the year, what is the resulting degree of confidence?

Answer. In general, if m out of n experts believe in some statement, then we assign, to this statement, the degree of certainty m/n. In our case, $m = 9$ and $n = 12$, so the degree of certainty is $9/12 = 0.75$.

4. Assuming that utility is proportional to the square root of money amount, would a person prefer \$1 without any condition or \$900 with probability 0.01?

Answer. According to decision theory, a person prefers an alternative with the largest utility.

- For the first alternative, the utility is $u_1 = \sqrt{1} = 1$ units.
- For the second alternative, the utility is

$$u_1 = 0.01 \cdot u(900) + 0.99 \cdot u(0) = 0.01 \cdot \sqrt{900} + 0.99 \cdot \sqrt{0} = 0.01 \cdot 30 + 0.99 \cdot 0 = 0.3.$$

The first utility is larger, so the person will prefer to get \$1 without any condition.

5. Why, for the smooth utility, moments are the most appropriate way of describing probability information when we need to make a decision?

Answer. There are many different representations of probability: cdf, probability density function (pdf), moments, etc. To decide which representation is most appropriate for decision making, we need to recall how probabilities are used in decision making.

How are probabilities used in decision making. According to decision theory, a rational decision maker selects an alternative for which the expected utility is the largest. So, to make this decision, we need to estimate, for each possible action, its expected utility. For each action, the expected utility is determined as follows:

$$u = p_1 \cdot u(x_1) + p_2 \cdot u(x_2) + \ldots + p_n \cdot u(x_n),$$

where:

- index $i = 1, \ldots, n$ describes possible outcomes of this action,
- p_i is the probability of the i-th outcome,
- x_i is the i-th outcome, and
- $u(x_i)$ is the utility of the i-th outcome.

Case when utility is a smooth function of the outcome. In many cases, small changes of outcome lead to equally small changes in utility, so the dependence $u(x)$ is smooth.

In such cases, it is usually convenient to expand this dependence in Taylor series and keep only several first terms in this expansion. This is, e.g., how functions like $\exp(x)$ and $\sin(x)$ are computed in the computers: we take into account that

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^N}{N!} + \ldots$$

and then estimate $\exp(x)$ as

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^N}{N!}.$$

So, we expand the dependence $u(x)$ in Taylor series and keep the first few terms in this expansion:

$$u(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \ldots + a_N \cdot x^N.$$

For this expression, the expected utility takes the following form:

$$u = p_1 \cdot (a_0 + a_1 \cdot x_1 + a_2 \cdot x_1^2 + \ldots + a_N \cdot x_1^N) +$$

$$p_2 \cdot (a_0 + a_1 \cdot x_2 + a_2 \cdot x_2^2 + \ldots + a_N \cdot x_2^N) +$$

$$\ldots +$$

$$p_n \cdot (a_0 + a_1 \cdot x_n + a_2 \cdot x_n^2 + \ldots + a_N \cdot x_n^N).$$

If we combine terms proportional to different powers of x_i, we get the following expression:

$$u = a_0 \cdot (p_1 + p_2 + \ldots + p_n) +$$
$$a_1 \cdot (p_1 \cdot x_1 + p_2 \cdot x_2 + \ldots + p_n \cdot x_n) +$$
$$a_2 \cdot (p_1 \cdot x_1^2 + p_2 \cdot x_2^2 + \ldots + p_n \cdot x_n^2) +$$
$$\ldots +$$
$$a_N \cdot (p_1 \cdot x_1^N + p_2 \cdot x_2^N + \ldots + p_n \cdot x_n^N).$$

The sum in the first parentheses is the sum of all the probabilities, i.e., 1. The sums in all other parentheses are expected values of some power of x_i. Such expected values are known as *moments*:

$$M_k = p_1 \cdot x_1^k + p_2 \cdot x_2^k + \ldots + p_n \cdot x_n^k.$$

In terms of moments, the above expression for the utility u takes the following form:

$$u = a_0 + a_1 \cdot M_1 + a_2 \cdot M_2 + \ldots + a_N \cdot M_N.$$

So: *in the case when utility is a smooth function of the outcome, to make a decision, it is desirable to known the moments.*

6a. If a value is 0 with probability 0.2, 10 with probability 0.3, and 20 with probability 0.5, what is the value F(4) of the corresponding cumulative distribution function? what is the value of F(30)?

Answer.

- Out of the three possible cases $x = 0$, $x = 10$, and $x = 20$, the only case when $x \le 4$ is when $x = 0$. The probability of this case is 0.2, so:

$$F(4) = \text{Prob}(x \le 4) = 0.2.$$

- The inequality $x \le 30$ happens in all three possible cases, so

$$F(30) = \text{Prob}(x \le 30) = 1.$$

6b. If a value is 0 with probability between 0.1 and 0.3, 10 with probability between 0.2 and 0.4, and 20 with the remaining probability, what is the interval value F(4) of the corresponding p-box? what is the value of F(30)?

Answer.

- Out of the three possible cases $x = 0$, $x = 10$, and $x = 20$, the only case when $x \le 4$ is when $x = 0$. The probability of this case is between 0.1 and 0.3, so:
$$F(4) = \text{Prob}(x \le 4) = [0.1, 0.3].$$

- The inequality $x \le 30$ happens in all three possible cases, so
$$F(30) = \text{Prob}(x \le 30) = 1.$$

7. If a value is 0 with probability 0.125, 10 with probability 0.125, 20 with probability 0.5, and 30 with probability 0.25, then how many binary questions do we have to ask, on average, to find the exact value?

Answer. The average number of binary questions is determined by the entropy $S = -\sum_i p_i \cdot \log_2(p_i)$, where p_i are the corresponding probabilities. In our case,

$$S = -0.125 \cdot \log_2(0.125) - 0.125 \cdot \log_2(0.125) - 0.5 \cdot \log_2(0.5) - 0.25 \cdot \log_2(0.25) =$$

$$-0.125 \cdot (-3) - 0.125 \cdot (-3) - 0.5 \cdot (-1) - 0.25 \cdot (-2) = 0.375 + 0.375 + 0.5 + 0.5 = 1.75.$$

8. Suppose that we have three different models of how fly epidemic spreads, and we use these models to estimate the probability that flu epidemic will reach El Paso in October. According to Models 1 and 3, this probability is 0.6, while Model 2 estimates the probability as 0.59. If the epidemic does reach El Paso in October, which model should we select is we use the Maximum Likelihood method? Which model should we select if in October, the epidemic will not yet reach El Paso?

Answer. Maximum likelihood means selecting the model for which the probability of observed events is the largest. So:

- If the epidemic does reach El Paso in October, then we should select either Model 1 or Model 3, since their probability of this event is larger: $0.6 > 0.59$.
- Models 1 and 3 predict the probability of the epidemic not reaching El Paso as $1 - 0.6 = 0.4$, while the Model 2 predicts $1 - 0.59 = 0.41$. So, if the epidemic does not reach El Paso in October, we should select Model 2, because for this model, the probability of this event is larger: $0.41 > 0.4$.

9. Suppose that we are use the Least Squares method for find the coefficients a and b of the linear dependence $y = a \cdot x + b$ between the quantities x and y. Based on the measurements $x_1 = 0$, $y_1 = 0$, $x_2 = 1$, $y_2 = 0$, $x_3 = 2$, $y_3 = 2$, what are the most probable values of a and b?

Answer. According to the least squares method, we have:

$$a = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\overline{x^2} - (\overline{x})^2},$$

and $b = \overline{y} - a \cdot \overline{x}$.

In this case,

$$\overline{x} = \frac{0+1+2}{3} = \frac{3}{3} = 1,$$

$$\overline{y} = \frac{0+0+2}{3} = \frac{2}{3},$$

$$\overline{x^2} = \frac{0+1+4}{3} = \frac{5}{3},$$

$$\overline{x \cdot y} = \frac{0 \cdot 0 + 1 \cdot 0 + 2 \cdot 2}{3} = \frac{4}{3},$$

so

$$a = \frac{\frac{4}{3} - 1 \cdot \frac{2}{3}}{\frac{5}{3} - 1^2} = \frac{\frac{2}{3}}{\frac{2}{3}} = 1,$$

and

$$b = \frac{2}{3} - 1 \cdot 1 = \frac{2}{3} - 1 = -\frac{1}{3}.$$

Thus, the dependence has the form

$$y = x - \frac{1}{3}.$$

10. If we know that the accuracy of x_1 is described by standard deviation $\sigma_1 = 0.01$, and the accuracy of x_2 is described by standard deviation $\sigma_2 = 0.03$, what is the accuracy of $y = 4x_1 + x_2$?

Answer. In general, for a linear combination $y = \sum c_i \cdot x_i$, the standard deviation is equal to $\sigma = \sqrt{\sum c_i^2 \cdot \sigma_i^2}$. In our case, $c_1 = 4$ and $c_2 = 1$, so:

$$\sigma = \sqrt{c_1^2 \cdot \sigma_1^2 + c_2^2 \cdot \sigma_2^2} = \sqrt{4^2 \cdot 0.01^2 + 1^2 \cdot 0.03^2} =$$

$$\sqrt{16 \cdot 0.0001 + 1 \cdot 0.0009} = \sqrt{0.0016 + 0.0009} = \sqrt{0.0025} = 0.05.$$