Decision Theory

- Shall I bring the umbrella today?
- I am looking for a house to buy. Shall I buy this one?
- Am I going to smoke the next cigarette?
- The court has to decide whether the defendant is guilty or not.
- Is this email spam or not?

We make decisions under uncertainty all the time!!

Decision Matrix

	Rain	No rain
Umbrella	Dry clothes Heavy Backpack True Positive	Dry clothes Heavy Backpack False Positive
No umbrella	Soaked clothes Light Backpack False Negative	Dry clothes Light backpack True Negative

How do we make decisions?

Given our belief, we take a possible action.

In this example the set of possible actions is {umbrella, no umbrella}. If we feel that P(rain) > P(no rain), we decide to bring an umbrella.

In this case our decision rule is based on a prior belief. To make better decision rules we should use information from the data as well.

SPAM Filters. We have two actions: α_1 for keep the mail and α_2 for delete as SPAM. There are two classes C_1: normal mail and C_2: SPAM

True Positive $\Rightarrow$ Correct Detection

False Negative $\Rightarrow$ Missed Detection

False Positive $\Rightarrow$ False Alarm

True Negative $\Rightarrow$ Correct Rejection

In a classification problem, the action is: Given the input **x**, assign observation to a class C_i. A **decision rule** $\alpha(x)$ is the action taken after observing x.

The **loss function**, denoted by λ, tells how costly each action is.

$\lambda(\alpha_i|C_j)$ is the loss incurred for taking action α_i if the true class is C_j .

An important loss function is zero-one-loss (0-1)

$\lambda(\alpha_i|C_j) = 0$ if $i = j$, and 1 if $i \neq j$

The expected loss or the **conditional risk** of taking action α_i, after observing **x**, is

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|C_j)P(C_j|\mathbf{x})$$

The total expected loss, also called **overall risk**, is

$$R(\alpha) = \int R(\alpha(\mathbf{x})|\mathbf{x}))p(\mathbf{x})d\mathbf{x}$$

We would like to find a decision rule $\alpha(x)$ that minimizes the overall risk (the minimum expected loss $R(\alpha_i|\mathbf{x})$).

This is **Bayes decision rule**. The overall risk $R(\alpha)$ for Bayes decision rule is called **Bayes risk**. This is the smallest possible overall risk.

Two Category Classification

The possible classes are C_1 and C_2 (or spam and not spam). The action α_1 corresponds to deciding if the true class is C_1 and $\alpha_2$corresponds to deciding that the true class is C_2.

$$R(\alpha_1|x) = \lambda_{11} P(C_1|\mathbf{x}) + \lambda_{12} P(C_2|\mathbf{x})$$

$$R(\alpha_2|x) = \lambda_{21} P(C_1|\mathbf{x}) + \lambda_{22} P(C_2|\mathbf{x})$$

What is Bayes decision rule?

Decide that the true class is C_1 if $R(\alpha_1|x) < R(\alpha_2|x)$ and C_2 otherwise.

$$R(\alpha_1|x) < R(\alpha_2|x)$$

$$(\lambda_{21} - \lambda_{11})P(C_1|\mathbf{x}) < (\lambda_{12} - \lambda_{22})P(C_2|\mathbf{x})$$

$$\frac{P(C_1|\mathbf{x})}{P(C_2|\mathbf{x})} < \frac{(\lambda_{21} - \lambda_{11})}{(\lambda_{12} - \lambda_{22})}$$

Here we assume that $\lambda_{21} > \lambda_{11}$, this is reasonable since the loss is greater when making a mistake.

In general we should pick the most probable case:

If $P(C_1|\mathbf{x}) > P(C_2|\mathbf{x})$ pick C_1

If $P(C_2|\mathbf{x}) > P(C_1|\mathbf{x})$ pick C_2