Adaptive Amusement Park

- An amusement park wants to adapt its games and rides to each visitor.
- Determine the gender of the visitor on entry by measuring some features. ω is either ω_1 (for male) or ω_2 (for female).
- From demographics: Probability $P(\omega = \omega_1) = P(\text{male}) = 0.45$. $P(\omega = \omega_2) = P(\text{female}) = 0.55$
- These are the **a priori** probabilities or **prior**s of male and female.

Guess the Gender

- A random person enters the park. What is the gender?
- **Decision Rule:**
  Decide ω_1 if $P(\omega_1) > P(\omega_2)$; else, decide ω_2.
- Females are more likely. Hence, guess is female for every visitor!
- A boring decision as the guess is same even if *know* males also will come.
- Why? Because we couldn't observe the person and have only the priors to guide us.

What if we can observe the height?

- Let x be the random variable indicating the height of the visitor, measured on entry.
- $p(x|\omega)$ gives the **class-conditional probability** density of the height x conditional on the class ω.
  It is the likelihood (or probability) of class ω generating the observation x.
- The *joint probability density* indicates when both events occur together. $p(x, \omega_i) = p(x|\omega_i)P(\omega_i) = P(\omega_i|x)p(x)$.
- **Bayes Formula:**
  $P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}$, where $p(x) = \sum_i p(x|\omega_i)P(\omega_i)$

What does it say?

- $p(w_i|x)$ gives the **a posteriori** or **posterior** probability that the state is ω_i given that the height is x.

- That is, the probability of male given an observed height of 135 cm.

- Bayes formula states: $$\textbf{posterior} = \frac{\textbf{likelihood} \times \textbf{prior}}{\textbf{evidence}}$$

- Bayes formula converts a priori probability $P(\omega_i)$ to a posteriori probability $P(\omega_i|x)$ with the help of the likelihood or conditional probability $p(x|\omega_i)$.

Modified Guess

- Observe the height. What is the modified decision?
- New decision rule: ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; else ω_2.

  Or, decide ω_1 if the **likelihood ratio** $\frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$.
- The LHS depends on the measurement. It should be greater than a threshold that depends on the priors.
- At 175 cm: likelihood for male is 0.7, and for female is 0.2. Decide male as $0.7/0.2 = 3.5 > 0.55/0.45 = 1.22$.
- At 135 cm: $p(x|\omega_2) = 0.8$ and $p(x|\omega_1) = 0.3$. Decide ω_2 as $0.3/0.8 < 1.22$.

Error in the Decision

- What's the probability of error, given an observation x?
- $P(error|x) = P(\omega_2|x)$ if decided on ω_1; else $P(\omega_1|x)$.
- For Bayes decision rule, $P(error|x) = \min\{P(\omega_1|x), P(\omega_2|x)\}$.
- Total error obtained by integrating over all x.

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error|x)p(x)dx$$

- Bayes decision rule minimizes $P(error|x)$ for every x. Therefore, it also minimizes the total error probability $P(error)$.

Risk in Decision

- Males get angry if misclassified as a female and damage property. Females also get angry but only sulk a little.

- Park owner wants of minimize his risk and would try to avoid males getting classified as females even at the cost of the reverse.

- Let α_i be an action i and $\lambda(\alpha_i|\omega_j) = \lambda_{ij}$ the risk in taking that action from state ω_j.

- Simple case: α_i means deciding ω_i.

Two-Category Classification

- Risk involved in each decision:

$$R(\alpha_1|\mathbf{x}) = \lambda_{11}P(\omega_1|\mathbf{x}) + \lambda_{12}P(\omega_2|\mathbf{x})$$
$$R(\alpha_2|\mathbf{x}) = \lambda_{21}P(\omega_1|\mathbf{x}) + \lambda_{22}P(\omega_2|\mathbf{x})$$

- Choose the decision with lower risk.
  Decide ω_1 if $(\lambda_{21} - \lambda_{11})P(\omega_1|\mathbf{x}) > (\lambda_{12} - \lambda_{22})P(\omega_2|\mathbf{x})$

- Using Bayes rule, this reduces to: Decide ω_1 if

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \frac{P(\omega_2)}{P(\omega_1)}$$

- Ordinarily, $\lambda_{12} > \lambda_{11}$ and $\lambda_{21} > \lambda_{22}$.

- As λ_{12} increases, the threshold increases. Makes it more difficult to decide ω_1.
- Makes sense as risk of misclassifying as ω_1 increases with λ_{12}.
- Risk moves the threshold in favour of the less risky action.
- In our example: $\lambda_{12} = 0.2$ (patient females) while $\lambda_{21} = 0.7$ (aggressive males).
- At 135 cm: $0.3/0.8 = 0.375 > 0.2*0.55/(0.7*0.45) = 0.35$. Hence, decide male!

The General Situation

- There are c states or **categories** $\{\omega_1, \omega_2, \ldots, \omega_c\}$.
- One of a actions $\{\alpha_1, \alpha_2, \ldots, \alpha_a\}$ is taken.
- Observation is a d-dimensional feature vector $\mathbf{x} = [x_1, x_2, \ldots, x_d]^\mathsf{T}$.
- $\lambda(\alpha_i|\omega_j)$: loss of taking action α_i given that the state is ω_j.
- Posteriors: $P(\omega_i|\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i)/\sum_i^c p(\mathbf{x}|\omega_i)P(\omega_i)$.
- Expected loss or **risk**: $R(\alpha_i|\mathbf{x}) = \sum_{j=1}^c \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x})$.
- Overall risk: $R = \int_\mathbf{x} R(\alpha(\mathbf{x})|\mathbf{x})p(\mathbf{x})d\mathbf{x}$.

Bayes Decision Rule under Risk

- Overall risk is minimized if the conditional risk $R(\alpha_i|\mathbf{x})$ is minimum for every $\mathbf{x}$.
- Bayes decision rule: Choose α_i that minimizes $R(\alpha_i|\mathbf{x})$
- Take action α_k (from $\alpha_1 \ldots \alpha_a$) where

$$k = \arg\min_i R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x})$$

.

- The minimum risk $R^*(\alpha_i|\mathbf{x})$ is called the **Bayes risk**.

Minimum Error-Rate Classification

- If we penalize wrong classifications equally, $\lambda_{ij} = 1 - \delta_{ij}$.
- Risk $R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda_{ij} P(\omega_j|\mathbf{x}) = 1 - P(\omega_i|\mathbf{x})$.
- $R(\alpha_i|\mathbf{x})$ is minimum for the decision i for which the posterior $P(\omega_i|\mathbf{x})$ is maximum.
- Same decision rule as the Bayes classifier.
- In the two-category case, if the loss for one action is greater than the other, the regions for that action will shrink.