Today:

- hypothesis testing / detection / classification
- MAP rule.

![Noisy Channel Diagram]

None of the problem:
- discrete $x$ and $y$
- statistics: hypothesis testing
- signal processing, communication: detection
- Machine learning: classification

**Model for noisy channel:**

$$p(y|x) \quad \text{or} \quad f_{y|x}(y|x)$$

- $y$ is discrete
- $x$ is continuous

Do we have a complete statistical model?

2) No, need the distribution of $x$

$$p_x(x) = P(x = x) \quad \text{for all } x \rightarrow \text{prior distribution.}$$

( In machine learning, $x$ is called $Y$, $Y$ is called $X$.)
detector:

maximum a posteriori rule: (MAP)

Choose \( x \) s.t. probability \( P_{\theta | x} (x | \theta) \) is largest.

\[
P_{\theta | x} (x | \theta) = \frac{P_{x | \theta} (x | \theta) \cdot P_{\theta} (\theta)}{P_{\theta} (\theta)}
\]

Bayes rule.

\[
P_{\theta} (\theta) = \sum_{\theta} P_{x | \theta} (x | \theta) \cdot P_{\theta} (\theta)
\]

maximum likelihood (ML):

Choose \( x \) s.t. \( P_{x | \theta} (x | \theta) \) is largest

Criterion: \( \text{Perre} = \Pr (\hat{X} \neq x) \)

Claim: MAP minimizes the probability of error.

ML minimizes the probability of error only when \( P(x \cdot 1) = P(x \cdot 0) \) (Or \( \theta \) in general, \( P_{\theta} (x) \) is uniform)

\[
P_{x | \theta} (x | \theta) : \begin{cases} 0.6 & \theta = 1 \\ 0.3 & \theta = 2 \\ 0.1 & \theta = 3 \\ \end{cases}
\]

We will pick "1".

\[
\Pr (\text{error} | \theta) = 1 - \frac{\sum_{\theta} a_{\theta}}{1} = 0.4
\]

* MAP minimizes \( \Pr (\text{error} | \theta) \) for all \( \theta \).

* \( \text{Perre} = \frac{1}{3} \sum_{\theta} \Pr (\text{error} | \theta) P_{\theta} (\theta) \)

\[= \text{MAP minimizes the overall Perre}\]
Under the uniform prior, $\text{MAP} = \text{ML}$.

$$P(x|y) = \frac{P(x|y)}{P(x)} \rightarrow \text{constant when } P_x(x) \text{ is uniform}$$

Bayes' rule

Moving to the continuous case

$$P(x|y) = \frac{f_x(x|y) \cdot P_x(x)}{f_x(x)}$$

Example

Binary detection / classification:

- $x = 0$ or $1$
- $P_x(0), P_x(1)$

Channel model:

- Conditional on $x = 0$, $Y \sim N(\mu_0, \sigma_0)$
- Conditional on $x = 1$, $Y \sim N(\mu_1, \sigma_1)$

E.g., sensor network

$$Y_i = h_i x + Z_i \quad i = 1, \ldots, n$$

$Z_i \sim N(0, \sigma^2)$, i.i.d.
\[ M_0 = 0 \in \mathbb{R}^n \quad (X = 0) \]
\[ M_i = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix} \quad (X = 1) \]
\[ K_0 = \sigma^2 I \in \mathbb{R}^{n \times n} \]
\[ K_i = \sigma^2 I \in \mathbb{R}^{n \times n} \]

Is this still a problem of the above form?

\[ Y_i = h_i X + \theta_i W + Z_i \]
\[ \sim N(0, \rho) \]

\[ W \text{ and } Z \text{ are independent} \]

\[ Y = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} X + \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} W + \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \]

\[ M_0 = 0, \quad M_i = h = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \]

\[ K_0 = K_i = \]

**total interference + noise:**

\[ Y = \theta W + Z \]

\[ K_Y = \mathbb{E}[Y Y^T] = \mathbb{E}[(\theta W + Z)(\theta W + Z)^T] = \mathbb{E}[\theta W (2W)^T] + \mathbb{E}[Z Z^T] \]
\[
\text{E} \left( W^2 gg^T \right) + \sigma^2 I
\]

\[
P gg^T + \sigma^2 I
\]

In sum: \( V = \text{int} + \text{noise} \)

\( K_v = \text{kin} + \text{noise} \)

\[
P gg^T \quad \text{full-rank}
\]

rank one matrix

---

**Note:** The model described above:

Machine learning: \( \rightarrow \) Gaussian discriminative model

\[
n \geq r \quad Y = X + Z
\]

\[
0 \text{ or } 1 \quad N(0, \sigma^2)
\]