Statistics

Hypothesis Testing & Confusion Matrix

1. Hypothesis Testing Setup

Hypothesis:

$$H_0 : \text{null hypothesis} \qquad \text{vs} \qquad H_1 : \text{alternative hypothesis}$$

In practice, we make a decision:

- **Reject** H_0, or
- **Fail to reject** H_0.

2. Four Possible Outcomes (Confusion Matrix View)

Machine learning confusion matrix

	Predict Positive	**Predict Negative**
Actual Positive	True Positive	False Negative
Actual Negative	False Positive	True Negative

- **False Positive (FP) Type I error**
- **False Negative (FN) Type II error**

Base knowledge

In $(22.10.5)$,

$$\text{statistical significance} = 1 - \alpha = 1 - P(\text{reject } H_0 \mid H_0 \text{ is true})$$

And in $(22.10.6)$,

$$\text{statistical power} = 1 - \beta = 1 - P(\text{fail to reject } H_0 \mid H_0 \text{ is false})$$

Thus,

$$\text{Type I error rate} = \alpha = P(\text{reject } H_0 \mid H_0 \text{ is true})$$

$$\text{Type II error rate} = \beta = P(\text{fail to reject } H_0 \mid H_0 \text{ is false})$$

Note:

- α also knows at Significant level.
- β also knows at Sensitivity.

Mapping into Machine learning Confusion Matrix

	Predict Positive	**Predict Negative**
Actual Positive	True Positive $=$ Correct decision	False Negative $= \beta$ $= P(\text{fail to reject } H_0 \mid H_0 \text{ is false})$
Actual Negative	False Positive $= \alpha$ $= P(\text{reject } H_0 \mid H_0 \text{ is true})$	True Negative $=$ Correct decision

Example

Hypothesis: Finding out new drug effect or Not on a group of patients.

$H_0(Null\ hypothesis)$: Drug has no effect vs H_1 : Drug has real effect exists

	Predict Drug Effect	**Predict Drug Not Effect**
Actual Effect	True Positive = Correct decision	False Negative $= \beta$ $= P(\text{fail to reject } H_0 \mid H_0 \text{ is false})$
Actual Not Effect	False Positive $= \alpha$ $= P(\text{reject } H_0 \mid H_0 \text{ is true})$	True Negative = Correct decision

Intepretation

- Type I error rate (FP):
  - False alarm - **Falsely** "discover" something.
    [Discover something exist, however, it does not.]
  - In ML terms: you predicted "positive" when the truth was negative
- Type II error rate (FN):
  - **Missing** "discover" something.
    [Skip a fact that is existed.]
  - In ML terms: you predicted “negative” when truth is positive.

Some reasons may lead to that and solvings

- Type I error rate (FP):
  - Random noise sample (Sample selection technique)
- Type II error rate (FN):
  - Sample size is small (So increase the sample size)
  - Results are noisy (So redesign experiment, or change the measurement)
  - The observed effect is weak (So increase the treatment, or the sample size)

Medical Screening

- Screening can falsely alarm healthy people or overlook real disease.
- Screening test would be cheap, easy to administer, and produce zero false negatives, if possible.
- Large sample.
  Hence, more Type I error (false positive) and less Type II error (false negative)

Vietnam COVID-19 Screening Example:

- **Hypothesis**: A person has COVID-19.
- **Null-Hypothesis** (H_0): A person does not have COVID-19.
- **Type I error (FP)**: The **test says a person has COVID-19** when they are **actually healthy**. Which leads to unnecessary isolation, anxiety, and resource use.
- **Type II error (FN):** The **test says a person is healthy** when they **actually have COVID-19**. Leads to undetected spread and delayed treatment.