# Eric Pitman Summer Workshop in Computational Science

## Cytotoxicity Classification Project Background

# Binary Classification

Separate elements in a data set into two groups, based on some property.

# Toxicity Information

| | Oxide nanoparticle | Toxicity | is.toxic |
|---|---|---|---|
| 2 | TiO2 | 1.74 | FALSE |
| 3 | SnO2 | 2.01 | FALSE |
| 4 | ZrO2 | 2.15 | FALSE |
| 5 | SiO2 | 2.2 | FALSE |
| 6 | Fe2O3 | 2.29 | TRUE |
| 7 | Al2O3 | 2.49 | TRUE |
| 8 | Cr2O3 | 2.51 | FALSE |

# Toxicity Information

- **Cytotoxicity** – found from experimental concentration of particles required to kill a certain number of cells

- **True value (is.toxic)** – whether a material is lethal or not in a worst-case scenario, from known health effects

# Binary Classification

- Goal: *automatically* determine if a material is toxic or not, based on some cytotoxicity threshold.

| | Oxide nanoparticle | Toxicity | is.toxic |
|---|---|---|---|
| 2 | TiO2 | 1.74 | FALSE |
| 3 | SnO2 | 2.01 | FALSE |
| 4 | ZrO2 | 2.15 | FALSE |
| 5 | SiO2 | 2.2 | FALSE |
| 6 | Fe2O3 | 2.29 | TRUE |
| 7 | Al2O3 | 2.49 | TRUE |
| 8 | Cr2O3 | 2.51 | FALSE |

Nontoxic

Toxic

# Binary Classification

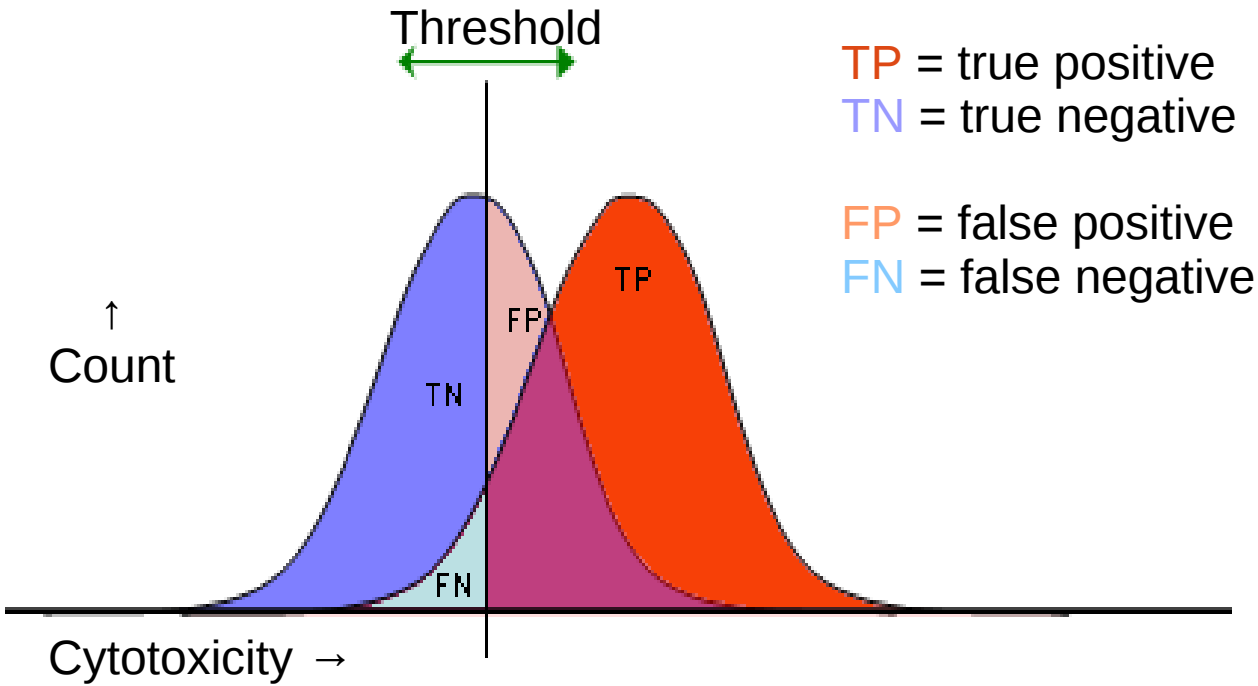
The best threshold value is unknown.

- We have a set of metal oxides known to be toxic or not, and their cytotoxicity values.
- We must use these data to find a threshold value that best classifies elements as toxic or non-toxic.

# Binary Classification

To perform classification, choose a cytotoxicity threshold.

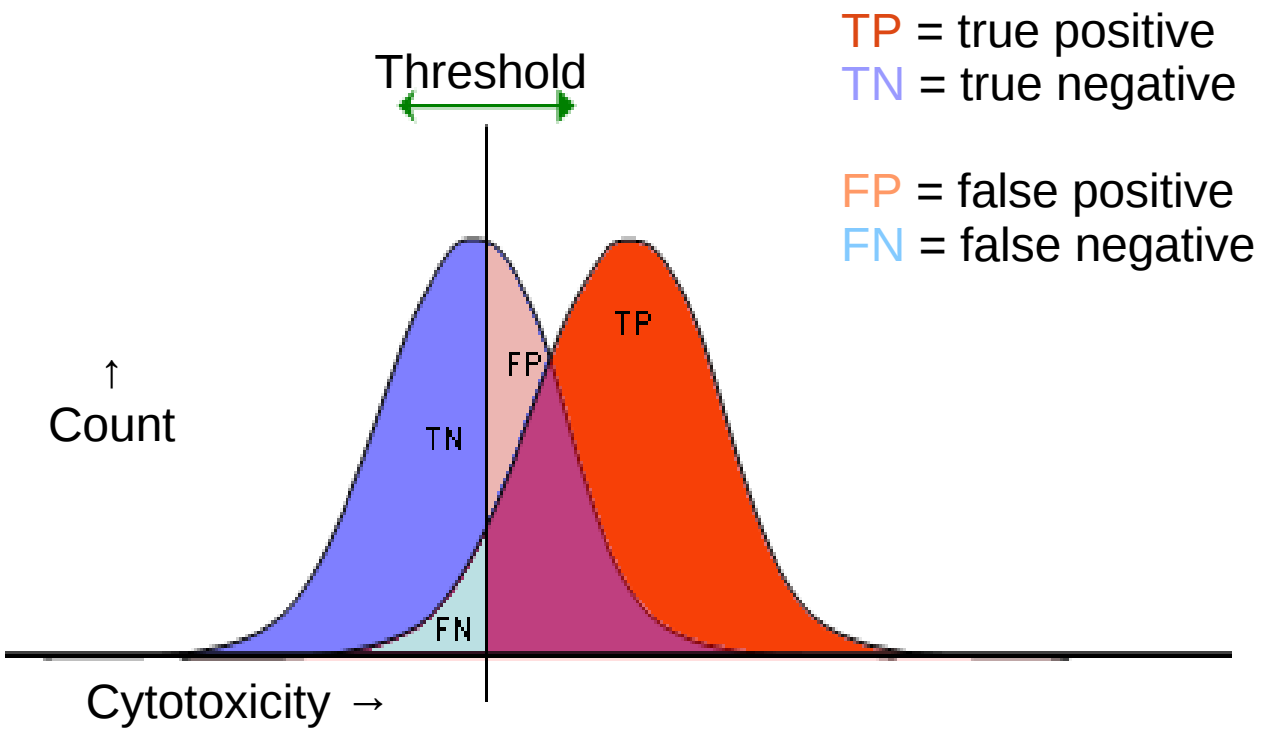
- Every value above the threshold is a positive result (toxic classification).
- Every value below the threshold is a negative result (non-toxic classification).



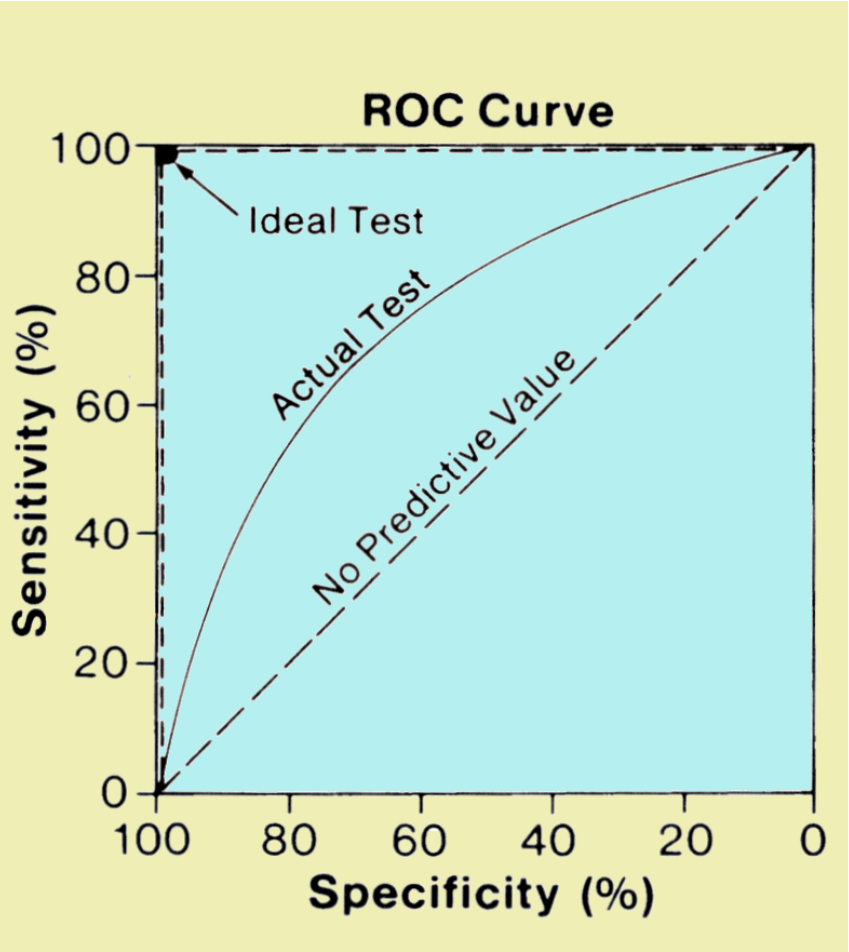
# Classifier Performance

Four possible outcomes:

- True and false positive (classified as toxic)
- True and false negative (classified as non-toxic)



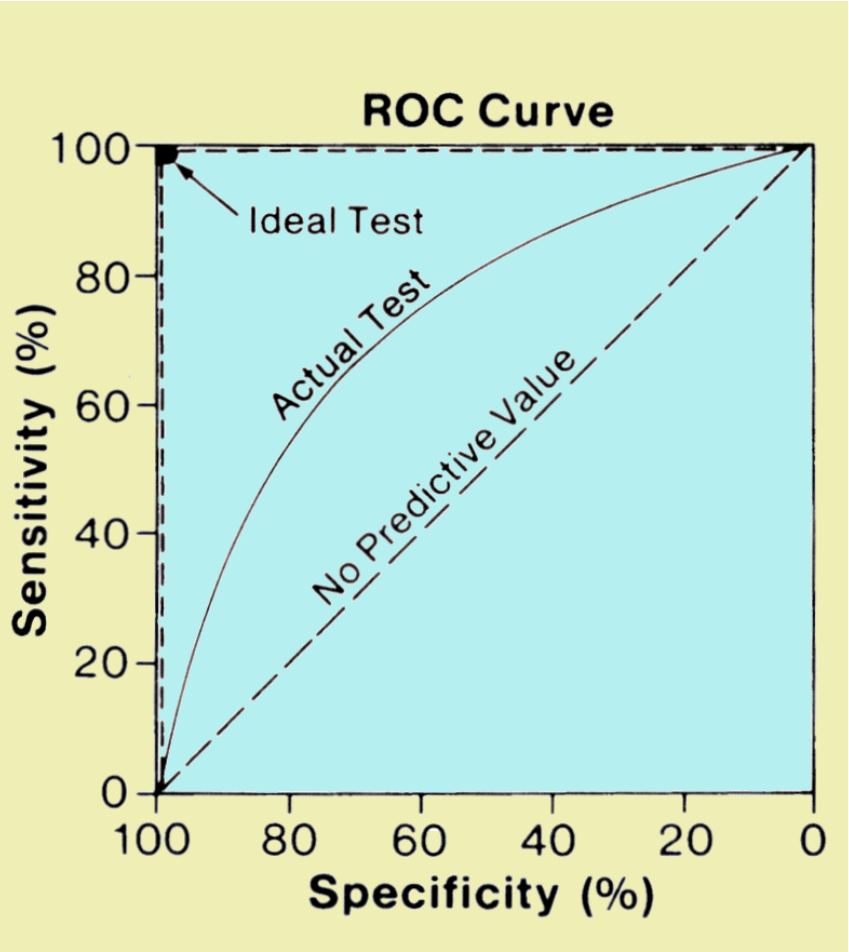
# Sensitivity and Specificity




- **Sensitivity** (true positive rate): The ability to correctly identify *toxic* substances

- **Specificity** (true negative rate): The ability to correctly identify *non-toxic* substances

# ROC Plot




- “Receiver Operating Characteristic” plot
- Used to evaluate binary classifier performance
- Compares true positive rate to false positive rate at a number of different thresholds