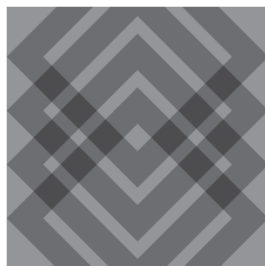


Eric Pitman Summer Workshop in Computational Science

Cytotoxicity Classification Project Background



CENTER FOR **COMPUTATIONAL RESEARCH**



University at Buffalo
The State University of New York

Binary Classification

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

From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!
Rolex w4tchs - \$100
Medicine (any kind) - \$50
Also low cost M0rgages
available.

Class 1

Spam

From: Alfred Ng
To: ang@cs.stanford.edu
Subject: Christmas dates?

Hey Andrew,
Was talking to Mom about plans
for Xmas. When do you get off
work. Meet Dec 22?
Alf

Non-spam

Class 2

Toxicity Information

1	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
1			
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

The diagram illustrates a binary classification process for oxide nanoparticles based on a toxicity threshold. The table lists eight nanoparticles with their respective toxicity values and whether they are classified as toxic (TRUE) or nontoxic (FALSE). A blue arrow labeled 'Nontoxic' points to the first four rows (TiO2, SnO2, ZrO2, SiO2), which all have toxicity values below 2.2 and are classified as FALSE. A red arrow labeled 'Toxic' points to the next two rows (Fe2O3, Al2O3), which have toxicity values above 2.2 and are classified as TRUE. The last row (Cr2O3) has a toxicity value of 2.51 but is classified as FALSE, which is an outlier from the general trend.

Binary Classification

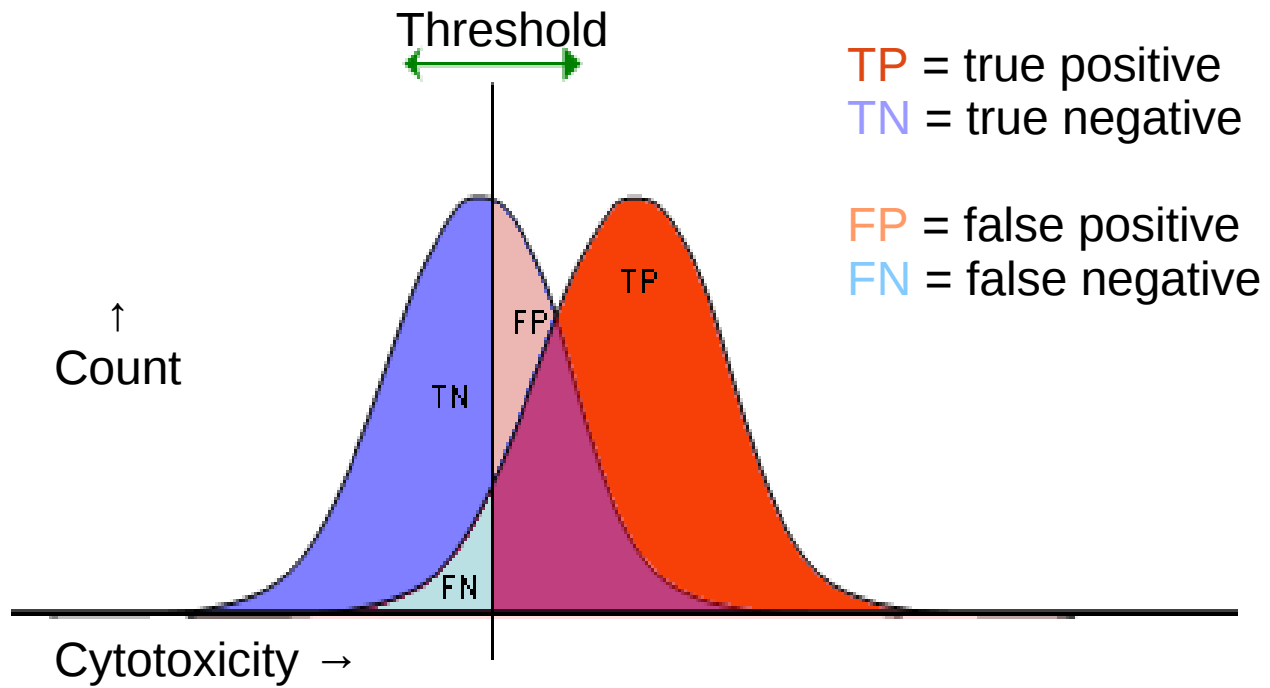
The best threshold value is unknown.

- 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.

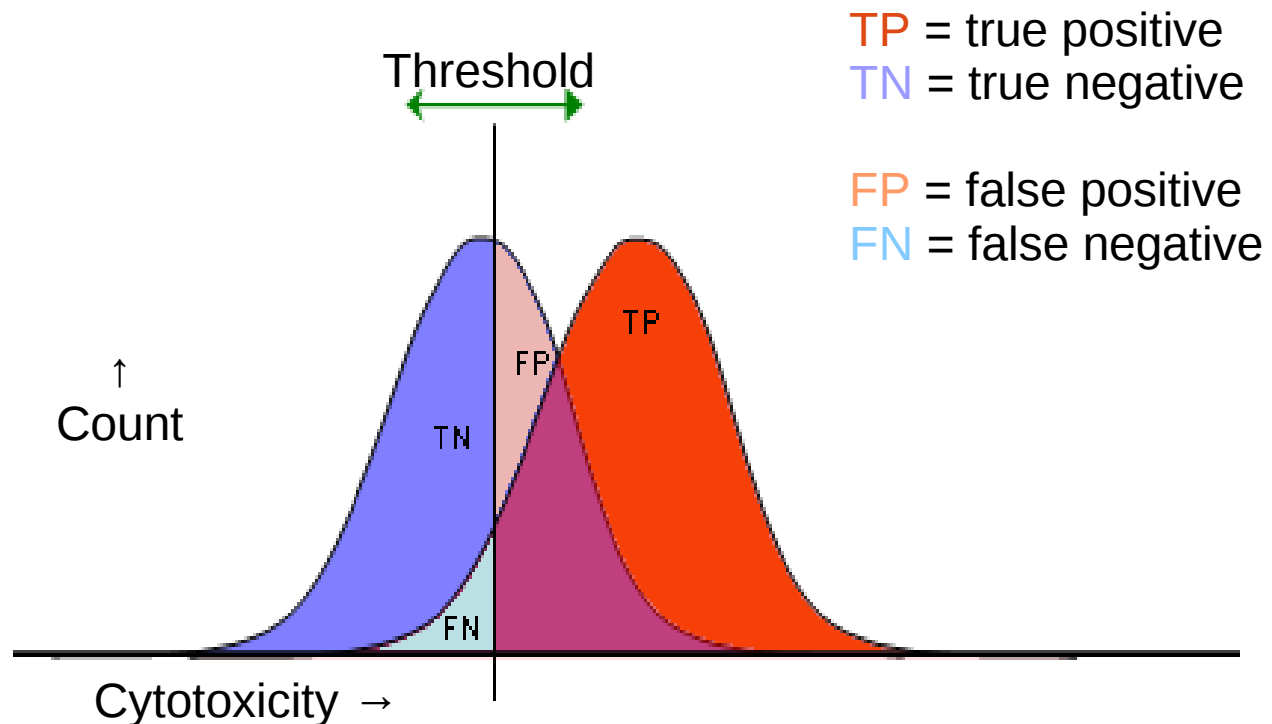
- Every value above the threshold is a positive result (toxic classification).
- Every value below the threshold is a negative result (non-toxic classification).



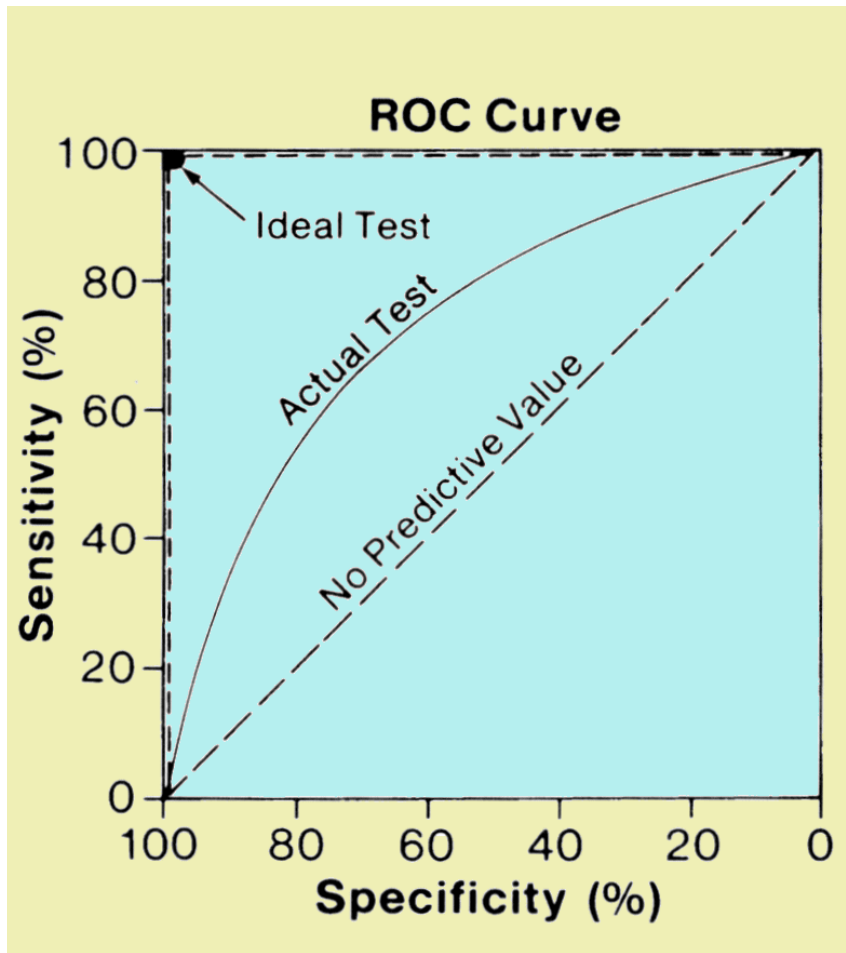
Classifier Performance

Four possible outcomes:

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

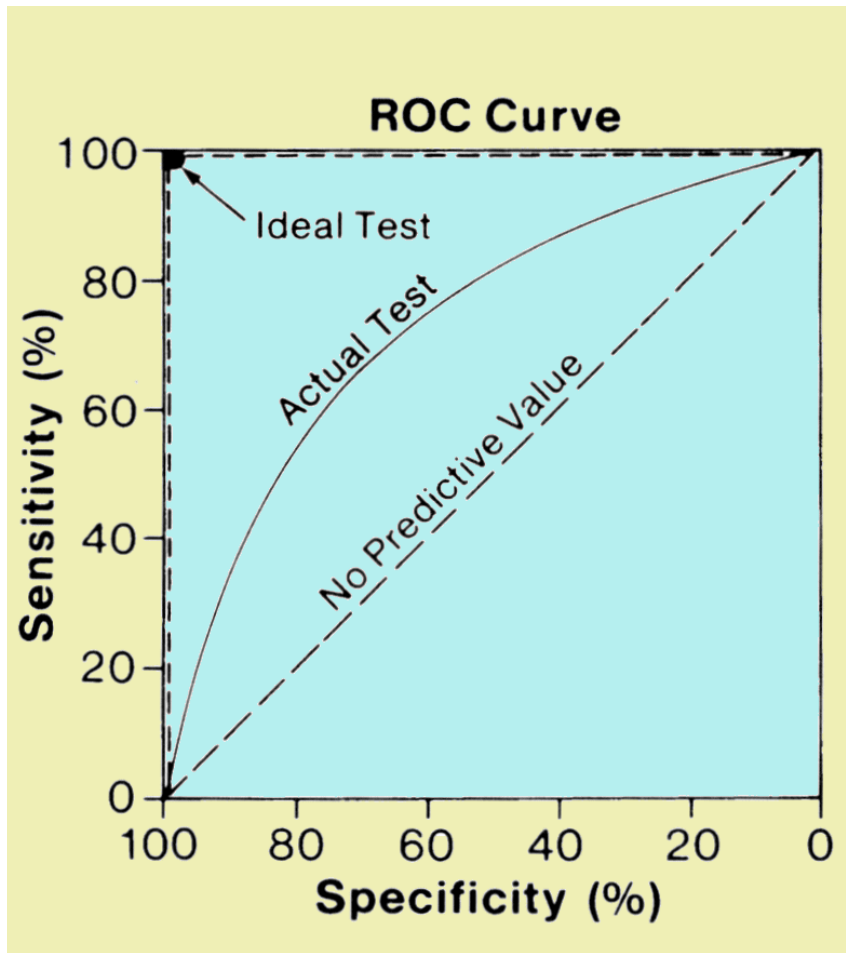


Sensitivity and Specificity



- **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