LDA (Linear Discriminant Analysis) in machine learning

Linear Discriminant Analysis (Normal Discriminant Analysis) is supervised classification problem that helps separate two or more classes by converting higher-dimensional data space into a lower-dimensional space. It is used to identify a linear combination of features that best separates classes within a dataset.
The problem

Key Assumptions of LDA
For LDA to perform effectively, certain assumptions are made:
-**Gaussian Distribution (Normal Distribution)**: the data in each class should follow the normal bell-shaped distribution
-**Equal Covariance Matrices**: All classes should have the same covariance structure.
-**Linear Separability**: The data should be separable using a straight line or plane.

Predictor = Feature = p

LDA when p=1

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{1}{2}\left(\frac{x-\mu_k}{\sigma_k}\right)^2}$$

Each class has it owns mean but all classes share the same variance

How LDA work?

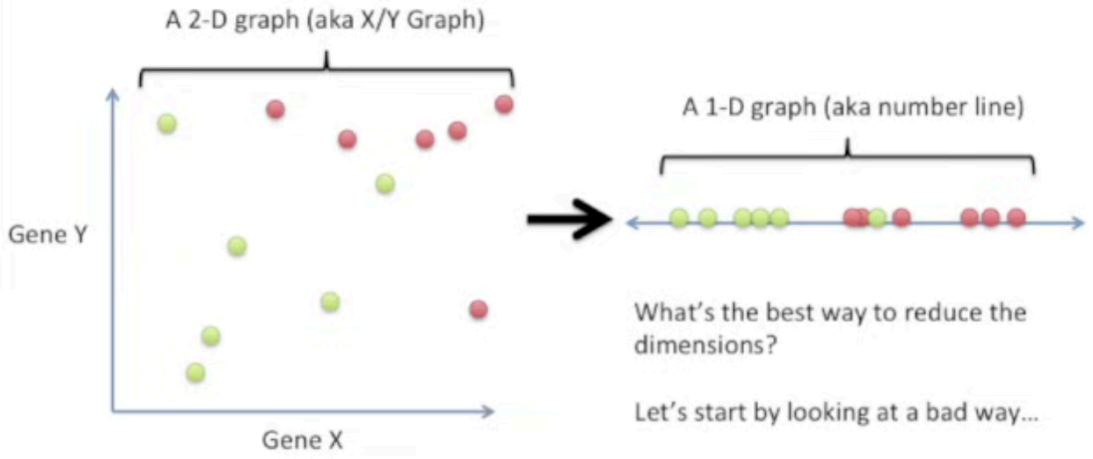
-We have got a cancer drug
-It work great for some people
-but it make other people feel worse

How do we decide who to give the drug to?

LDA is like PCA, but it focuses on maximizing the seperatibility among known categories.

A super simple example

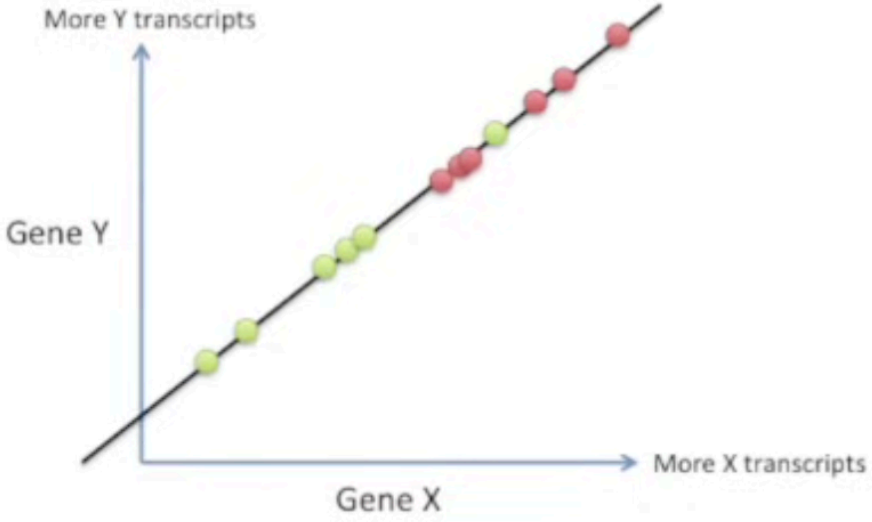
Reducing a 2-D graph to a 1-D graph



What's the best way to reduce the dimensions?

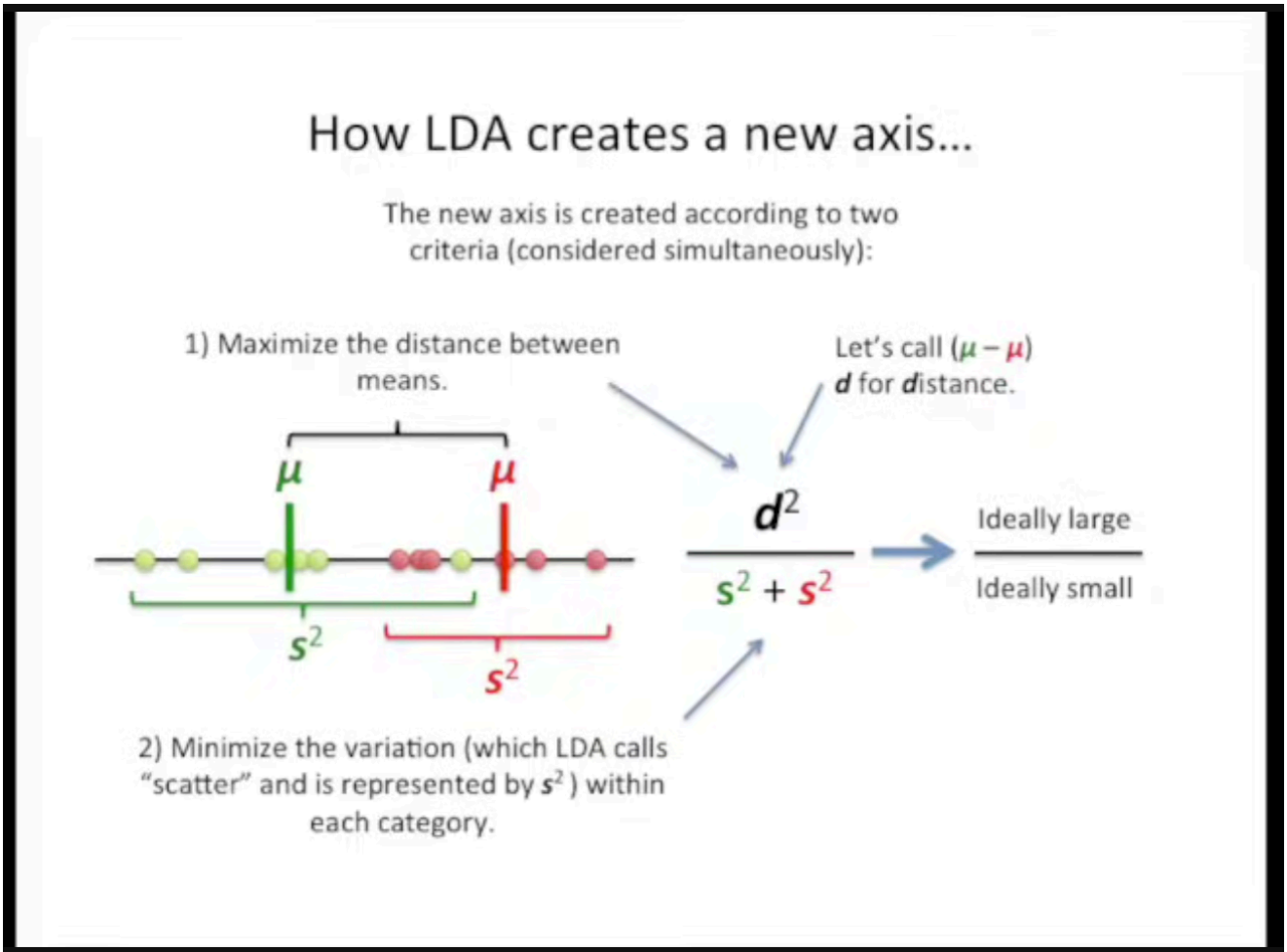
Let's start by looking at a bad way...

Reducing a 2-D graph to a 1-D graph with LDA



LDA uses both genes to create a new axis...

...and projects the data onto this new axis in a way to maximize the separation of the two categories.



🔍 LDA vs QDA (quick comparison)

Feature	LDA	QDA
Covariance	Shared	Class-specific
Boundary	Linear	Quadratic
Flexibility	Low	High
Overfitting risk	Low	High
Data needed	Less	More
Bias–Variance	Higher bias	Higher variance

Bias–Variance intuition 💡

- **LDA** → more bias, less variance
- **QDA** → less bias, more variance

If your dataset is small → **LDA usually wins**

If dataset is large & complex → **QDA can shine**