LINEAR DISCRIMINANT ANALYSIS

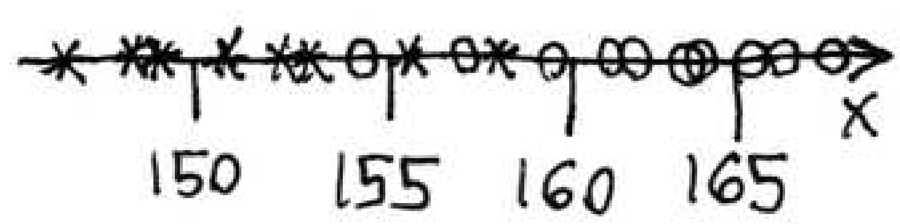
LDA is a method to assign a "class" (ie, label $k = 1, 2, \cdots, K$) to a datapoint x comprising n variables, given a "training set" of other datapoints with known labels. The goal is to correctly predict labels of datapoints — it is one of the earliest "machine learning" algorithms.

We now adapt examples from C.R. Rao's foundational 1948 paper on LDA, keeping their racial & eugenic problems fully in mind...

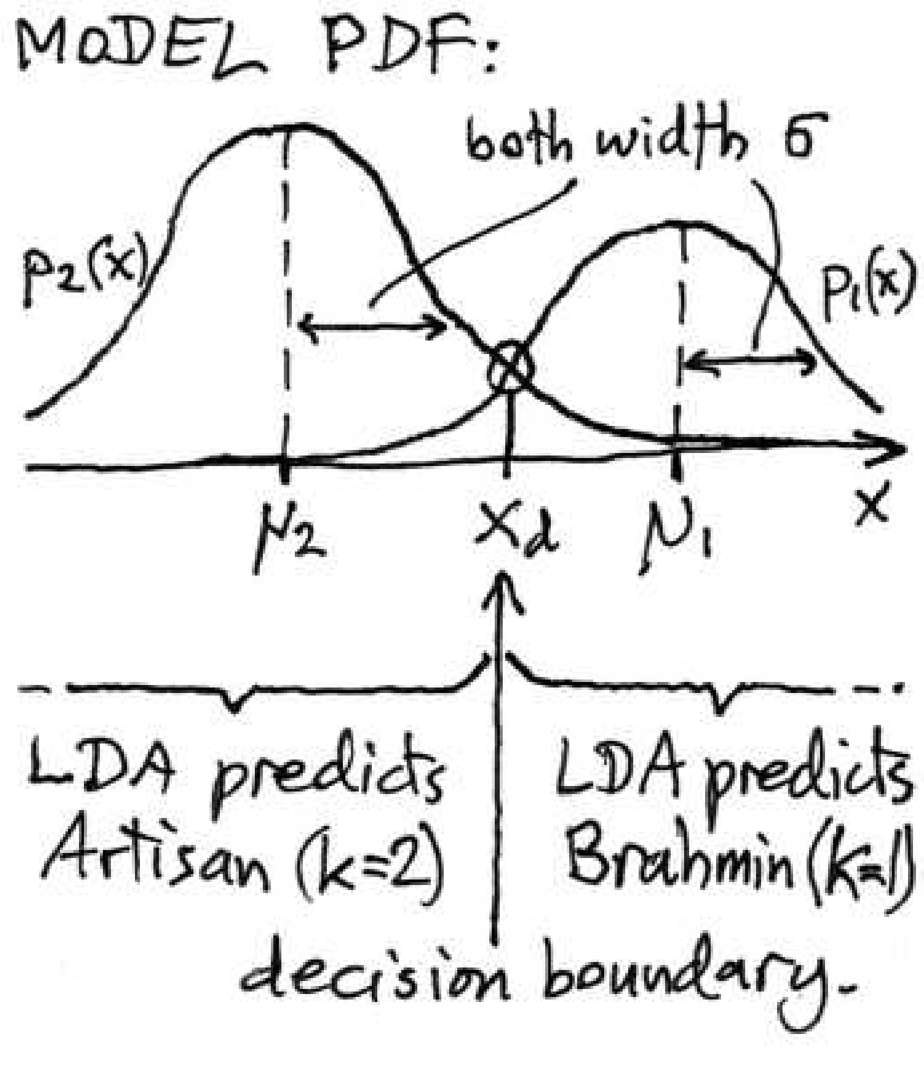
$\boxed{1D}$ We start in $n=1$ dimensions:

Let x be a human's stature (height) in cm. From such biometrics, Rao (working with Mahalanobis) wished to predict caste, ie assign labels $k=1$ (Brahmin), $k=2$ (Artisan), etc. For now, let's stick to those two classes ($K=2$). The training set is heights of random citizens, of known caste: two types of point (o or *) scattered on the x axis as shown to the right.

TRAINING DATA:
o Brahmin ($k=1$)
* Artisan ($k=2$)



MODEL PDF:



LDA models the training data by a Gaussian (normal) "probability density function" (pdf) for x in each class, with the same variance σ^2, but different means μ_1, μ_2, and different "masses" (areas) π_1, π_2. A pdf is simply a graph showing the expected distribution of x. The previous figure sketches the fitted model pdf: two "bell curves" of different heights. Here's the formula:

$$p_k(x) = \frac{\pi_k}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu_k)^2}{2\sigma^2}} \qquad k=1,2 \quad x \in \mathbb{R}$$

The point is that the four parameters $(\mu_1, \mu_2, \sigma, \pi_1)$ are easy to estimate from training data:

μ_1 = mean height of Brahmins

μ_2 = mean height of Artisans

π_1 = fraction that are Brahmin $= 1 - \pi_2$

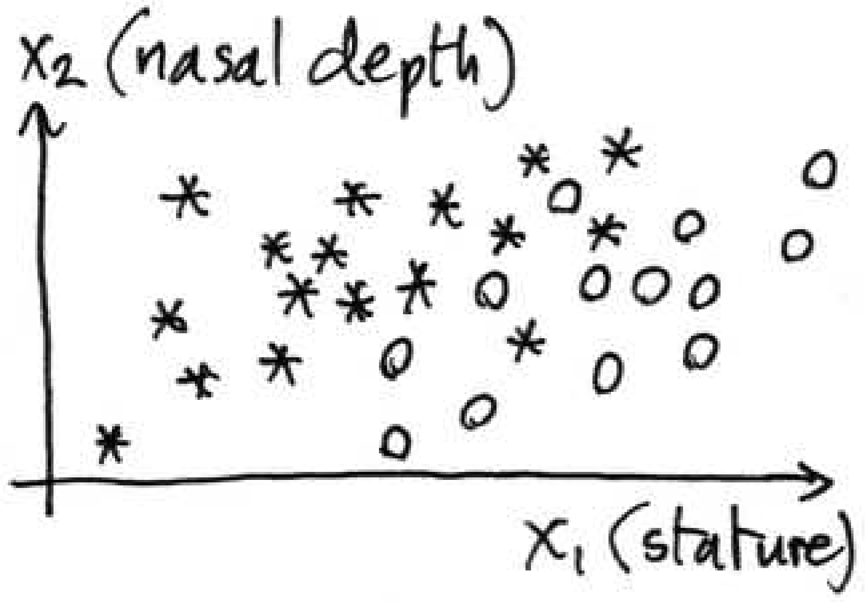
σ^2 = mean square deviation of each height from its respective class mean (μ_1 or μ_2).

Now let's predict! Given a new x to classify, LDA simply picks the most likely k conditioned on this x, which is the same as asking if $p_1(x)$ or $p_2(x)$ is the larger. The "decision boundary" x_d is where $p_1(x_d) = p_2(x_d)$, ie, where the two curves cross (see figure). Any $x > x_d$ is predicted Brahmin, any $x < x_d$ Artisan.

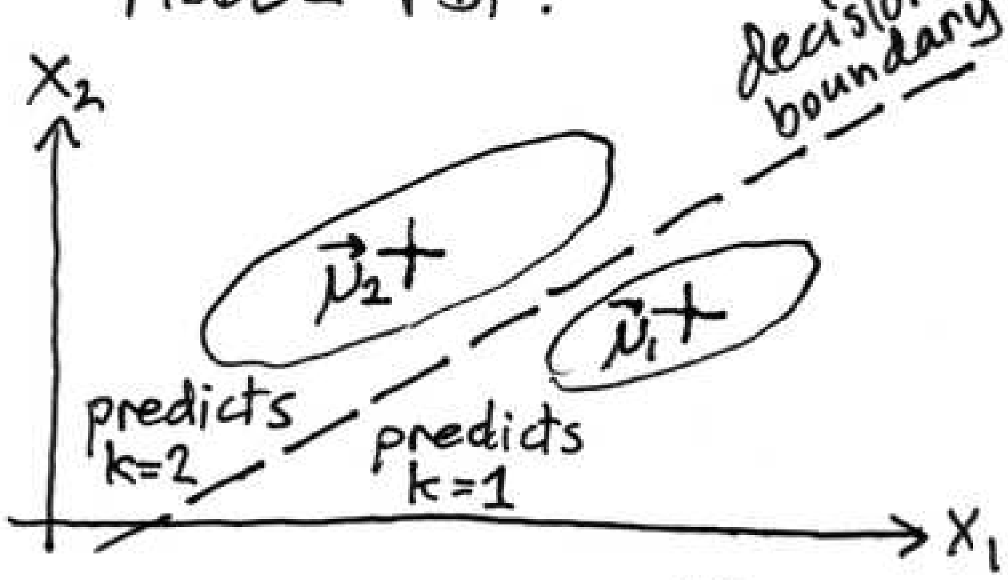
- Note, the larger the separation $D = \frac{|\mu_2 - \mu_1|}{\sigma}$ (called "Mahalanobis distance"), the higher expected prediction accuracy. Here, $D \approx 2$. If $D < 1$ you may as well just toss a coin!

2D & higher Since more variables are better (right?), LDA is more powerful in $n=2$, or higher, dimensions. Rao considered $\vec{x} = (x_1, x_2)$ where x_1 = stature, x_2 = nasal depth. Now each citizen is a datapoint in 2D space:

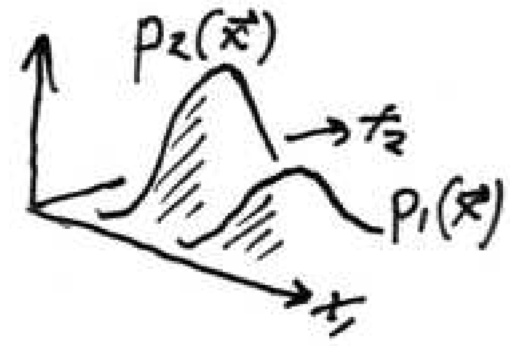
TRAINING DATA:



MODEL PDF:

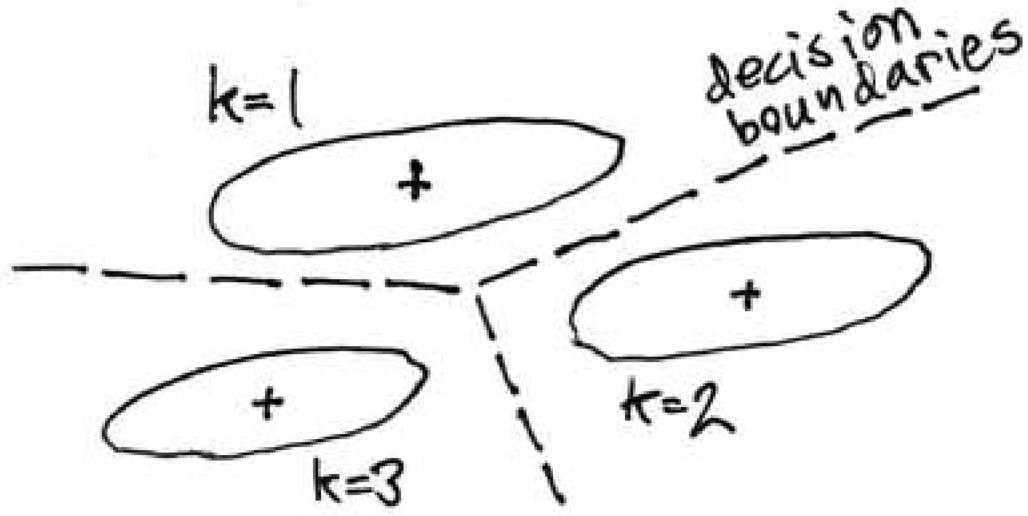


3d graph:

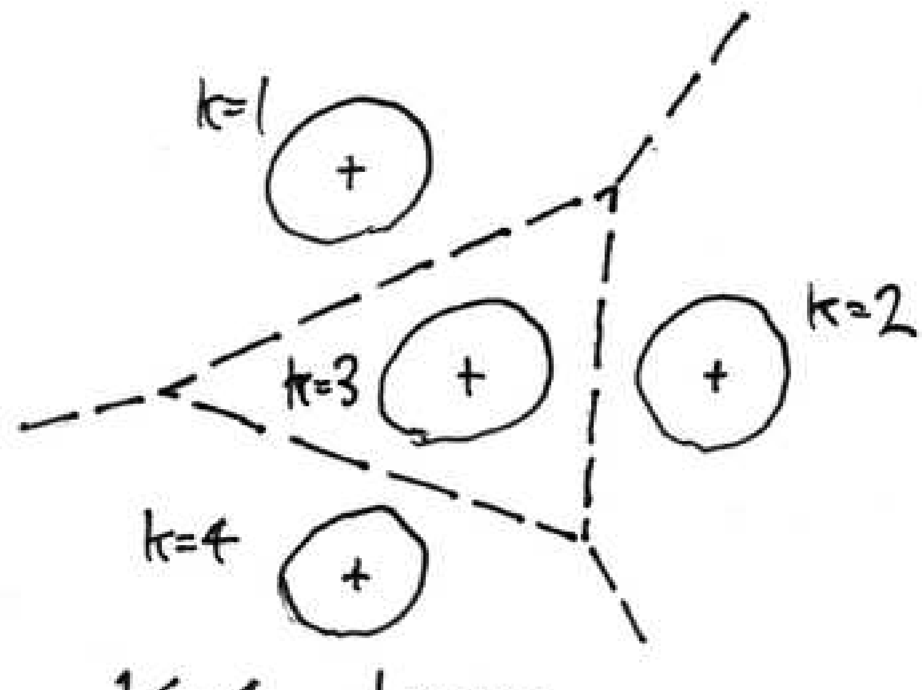


Again, LDA models the training data (by a mixture of "multivariate" Gaussians with the same <u>covariance</u>, as shown by the skew ellipses above), then chooses a decision boundary by setting $p_1(\vec{x}) = p_2(\vec{x})$: the result is a <u>hyperplane</u> (in 2D, a line), hence "linear discriminant". The covariance is usually found via PCA (see Fig. 37).

K>2 classes LDA also extends to more than 2 class labels. Setting pairs of densities $p_j(\vec{x}) = p_k(\vec{x})$ equal leads to various touching <u>pieces</u> of hyperplanes, fracturing data space in a similar fashion to a "Voronoi tesselation". Here's a sketch of some examples in 2D:



K=3 classes.



K=4 classes.

Notes on LDA:

- For the full formulae, see the References.
- LDA is easy to use, but makes <u>strong</u> assumptions, rarely true:
  - classes are discrete, stable, cover all possibilities.
  - all pdfs are Gaussians, of same covariance.
- <u>Non</u>linear decision boundaries are more flexible, eg, via support vector machines (SVM) or neural networks (NN), but need <u>more</u> training data & computer time. In the last decade, <u>convolutional NNs</u> have radically improved accuracy in difficult image classification & recognition problems.
- LDA (and all the fancier above classifiers) reproduce the discrimination in the training data, which is usually selected by a human, or other algorithms... <u>bias in, bias out</u>. (BIBO).
- Rao's work did/does not prove that caste is racial, and ignores crucial social & environmental factors. If "Brahmins" and "Artisans" were treated & nourished equally, would their height pdfs tend to become the same?