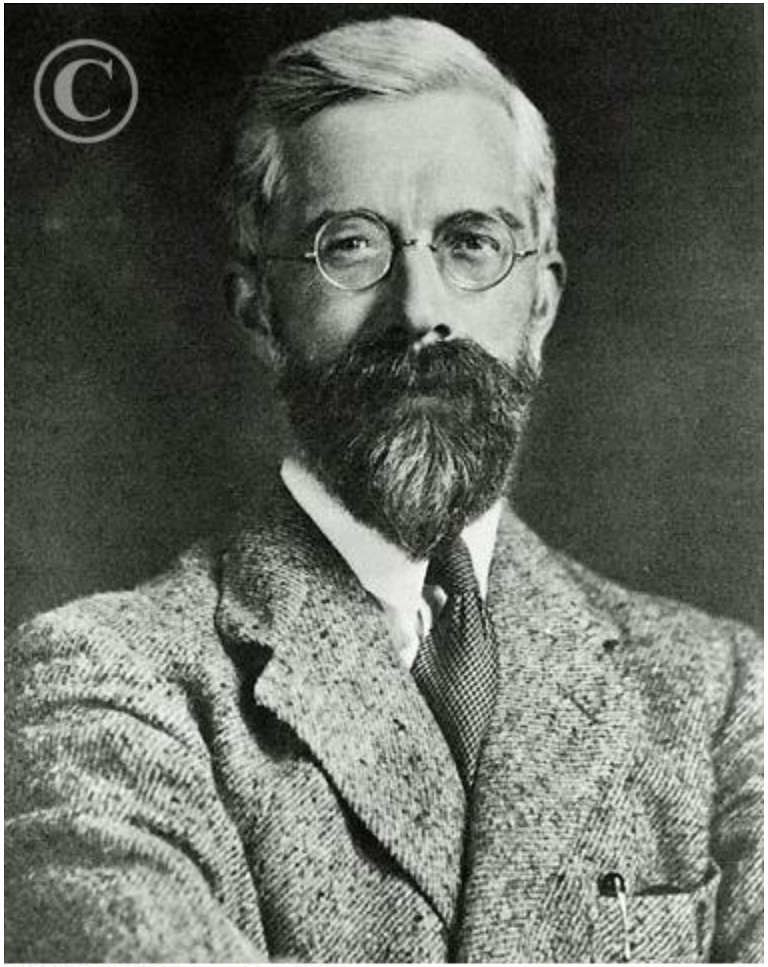
Linear discriminative analysis

By xudong

02/01/2012

Ronald Fisher

- 17 February 1890 – 29 July 1962
- English statistician, evolutionary biologist, eugenicist and geneticist
- Fisher's exact test and Fisher's equation
- "a genius who almost single-handedly created the foundations for modern statistical science"
- "the greatest biologist since Darwin".



Fisher's exact test

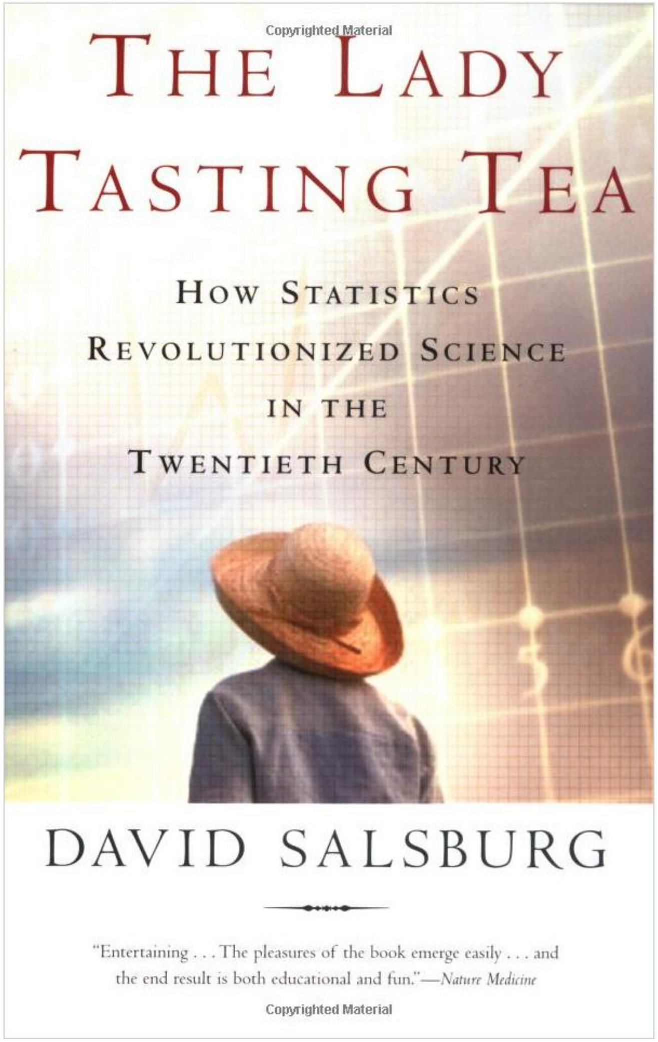
	Milk First	Tea First	
Milk First	5	0	5
Tea First	1	4	5
	6	4	10

Null hypothesis: **The lady don't have the ability**

$$P = \frac{(R_1!\,R_2!\,R_3!\ldots R_m!)(C_1!\,C_2!\,C_3!\ldots C_m!)}{N!\prod_{i,j} a_{\mathrm{ij}}}$$

$$P = \frac{4!\,6!\,5!\,5!}{10!\,5!\,1!\,0!\,4!} = 0.0238$$

The odd is too small to accept the Null hypothesis



C. R. Rao

Knowledge is what we know
Also, what we know we do not know.
We discover what we do not know
Essentially by what we know.

Thus knowledge expands.

With more knowledge we come to know
More of what we do not know.

Thus knowledge expands endlessly.

* * *

All knowledge is, in final analysis, history.
All sciences are, in the abstract, mathematics.
All judgements are, in their rationale, statistics.

Statistics and truth: put chance to work

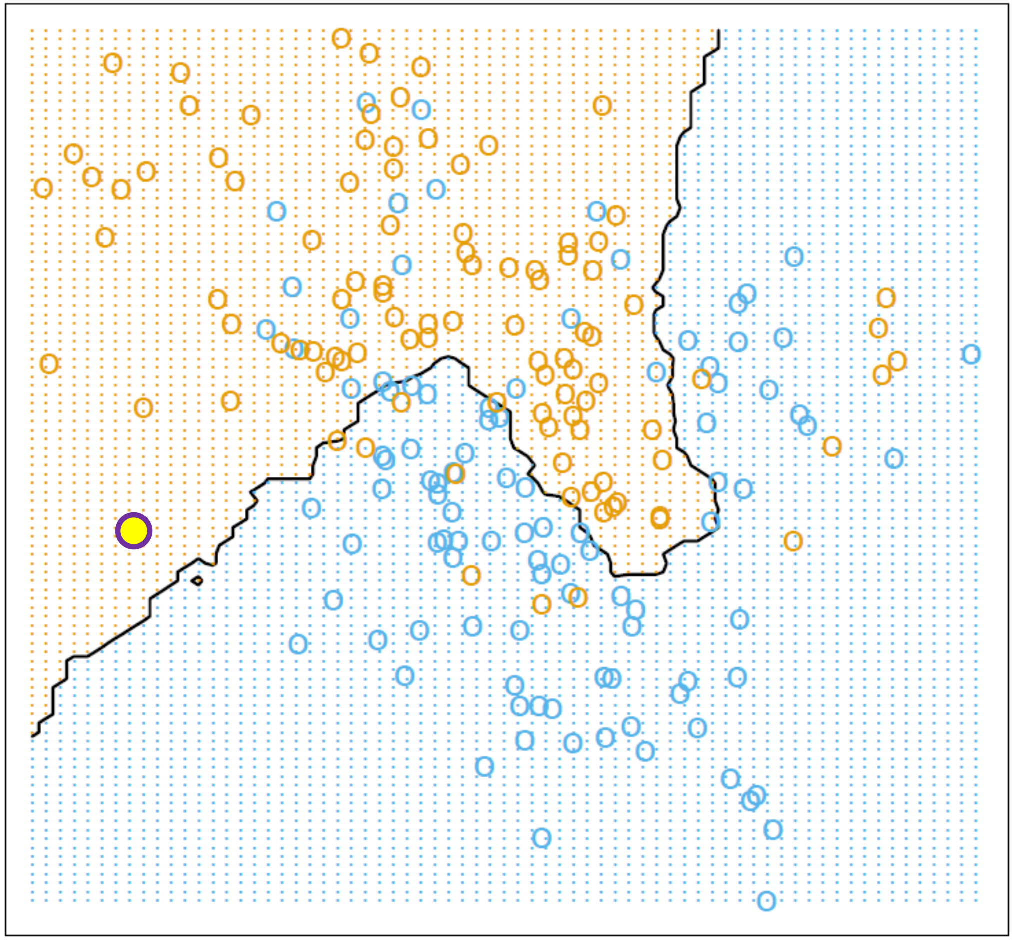
outline

- LDA for classification
  - Derivation: Bayesian way
  - Example: Altman Z-score
- LDA for supervised dimension reduction
  - Derivation: maximize $R = \sigma_{\text{between}} / \sigma_{\text{wihtin}}$
- LDA for clustering
  - Discriminative clustering

outline

- LDA for classification
  - Derivation: Bayesian way
  - Example: Altman Z-score
- LDA for supervised dimension reduction
  - Derivation: maximize $R = \sigma_{\text{between}} / \sigma_{\text{wihtin}}$

Classification



- Which class?

Bayesian method for classification

- Compute $P(k|x)$
  - $P(k|x) = \frac{P(x|k)P(k)}{P(x)}$ (Bayesian rule)
  - $P(k|x) \propto P(x|k)P(k)$

- $c = \underset{k}{\operatorname{argmax}} P(k|x)$

Classification with Gaussian hypothesis

- $P(k|x) \propto P(x|k)P(k)$
- $P(x|k) = N(\mu_k, \Sigma_k)$
  - $N(\mu, \Sigma) = \frac{1}{(2\pi)^{m/2}\det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$
  - $\mu = \frac{\sum x_i}{N}$
  - $\Sigma = \frac{\sum (x_i - \mu)(x_i - \mu)^T}{N-1}$
- $c = \underset{k}{\operatorname{argmax}}\, P(k|x)$

Special case (1) - QDA

- Two classes
  1) $P(1|x) > P(2|x)$
  2) $\log P(1|x) > \log P(2|x)$
  3) $\log P(x|1)P(1) > \log P(x|2)P(2)$
  4) $(x-\mu_1)^T {\Sigma_1}^{-1}(x-\mu_1) + \log\det(\Sigma_1) - (x-\mu_2)^T {\Sigma_2}^{-1}(x-\mu_2) - \log\det(\Sigma_2) > \quad \log P(1) - \log P(2)$
- Quadratic discriminant analysis

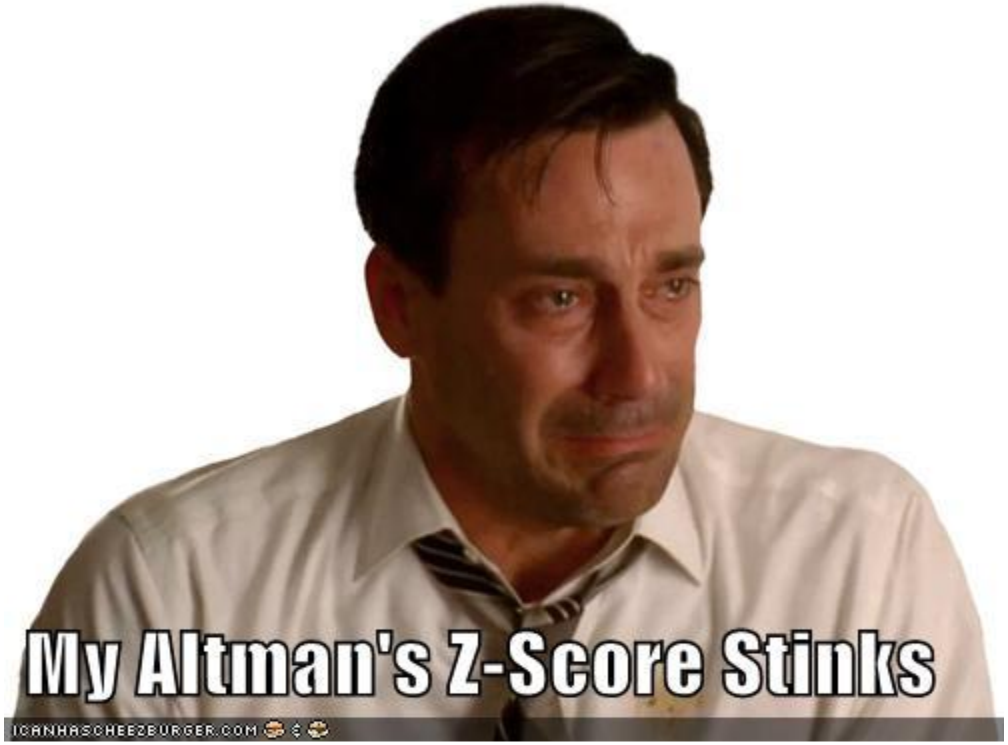
Special case (2) - LDA

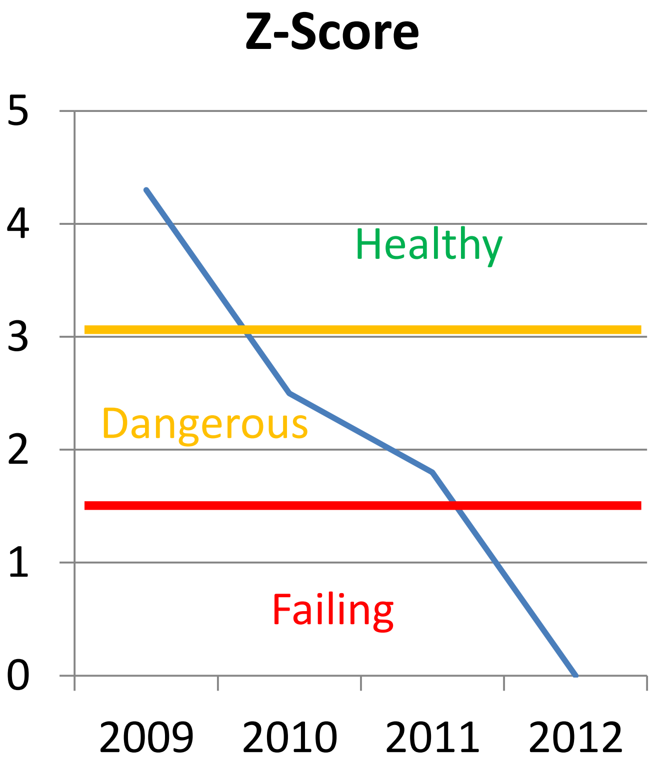
- Two classes with identical covariances
  - $\Sigma_1 = \Sigma_2 = \Sigma$
    - $(x-\mu_1)^T {\Sigma_1}^{-1}(x-\mu_1) + \text{logdet}(\Sigma_1) - (x-\mu_2)^T {\Sigma_2}^{-1}(x-\mu_2) + \text{logdet}(\Sigma_2) > \log P(1) - \log P(2)$
    - $(x-\mu_1)^T \Sigma^{-1}(x-\mu_1) - (x-\mu_2)^T \Sigma^{-1}(x-\mu_2) > \log P(1) - \log P(2)$
  - $x^T \Sigma^{-1}(\mu_1 - \mu_2) > \frac{1}{2}(\log P(1) - \log P(2) - {\mu_1}^T \Sigma^{-1}\mu_1 + {\mu_2}^T \Sigma^{-1}\mu_2)$
- Linear discriminative classifier
  - $w^T x > c$
    - $w = \Sigma^{-1}(\mu_1 - \mu_2)$
    - $c = \frac{1}{2}(\log P(1) - \log P(2) - {\mu_1}^T \Sigma^{-1}\mu_1 + {\mu_2}^T \Sigma^{-1}\mu_2)$

outline

- LDA for classification
  - Derivation: Bayesian way
  - Example: Altman Z-score
- LDA for supervised dimension reduction
  - Derivation: maximize $R = \sigma_{\text{between}} / \sigma_{\text{wihtin}}$
- LDA for clustering
  - Discriminative clustering

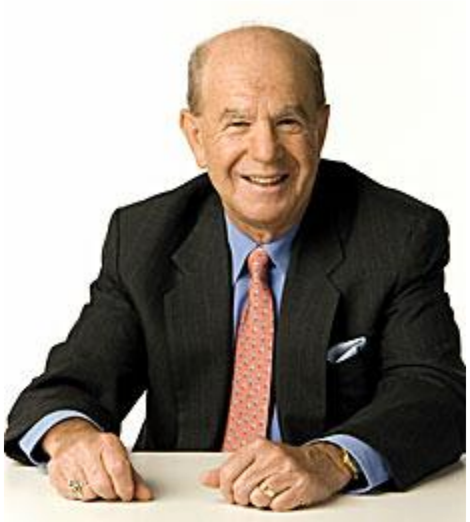
Altman Z-Score (1)





Altman Z-Score (2)

- predicting bankruptcy
- published in 1968
- A leading model in practical applications
  - 80-90% accurate
  - Type II error rate 15-20%



Edward I. Altman

http://en.wikipedia.org/wiki/Z-Score_Financial_Analysis_Tool

Altman Z-Score (3)

- Z-score estimated for private firms
  - T_1 = (Current Assets-Current Liabilities) / Total Assets
  - T_2 = Retained Earnings / Total Assets
  - T_3 = Earnings Before Interest and Taxes / Total Assets
  - T_4 = Book Value of Equity / Total Liabilities
  - T_5 = Sales/ Total Assets
- Z' Score Bankruptcy Model:
  - $Z' = 0.717T_1 + 0.847T_2 + 3.107T_3 + 0.420T_4 + 0.998T_5$
  - $w^T x>c$
  - $w = \Sigma^{-1}(\mu_1 - \mu_2)$
- Zones of Discrimination:
  - $Z' > 2.9$ - Safe
  - $1.23 < Z' < 2.9$ - Dangerous
  - $Z' < 1.23$ - Failing

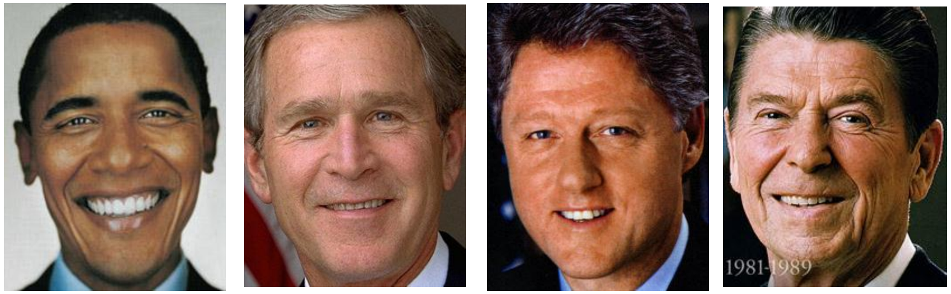
http://en.wikipedia.org/wiki/Z-Score_Financial_Analysis_Tool

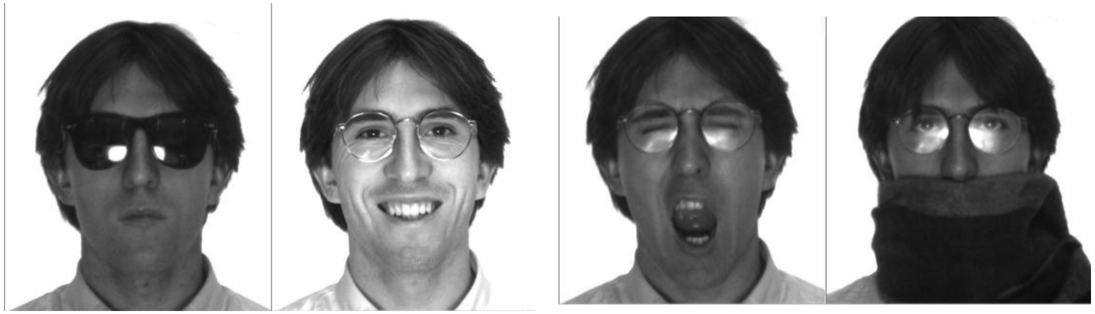
outline

- LDA for classification
  - Derivation: Bayesian way
  - Example: Altman Z-score
- LDA for supervised dimension reduction
  - Derivation: $\text{maximize}\ R = \sigma_{\text{between}}/\sigma_{\text{wihtin}}$
  - Example: Regression in subspace
- LDA for clustering
  - Discriminative clustering

Supervised dimension reduction(linear)

- Two factors contribute to the appearance
  - identity
  - Expression, illumination, occlusion, pose, age etc.
- project the original vector into a subspace with a fat matrix (row<col)
  - Keep the information of identity
  - Remove irrelevant information





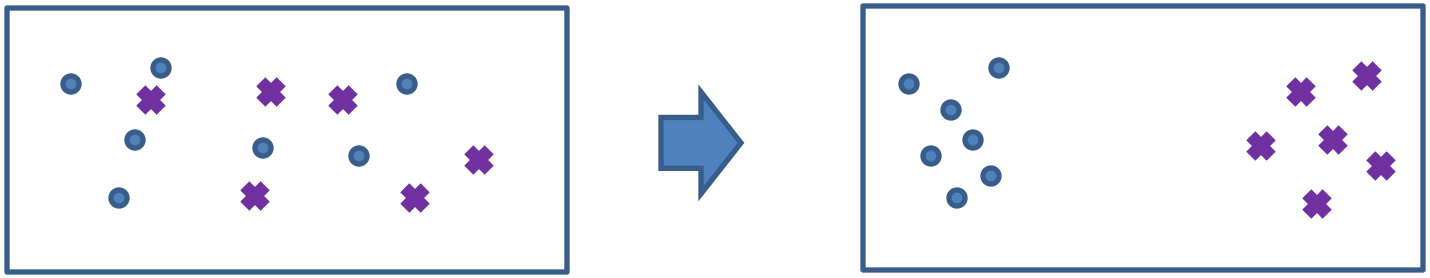
Fisher's linear discriminant(1)

- Each sample is decomposed as $x = \mu + \epsilon$
  - μ is the center of the corresponding class
- Find one direction w such as

- Minimize within class variance
  - $\min \quad w^T \Sigma_{\text{between}} w$
  - $\Sigma_{\text{between}} = \frac{\sum_{i=1}^{N} (\mu_i - \mu)(\mu_i - \mu)^T}{N-1}$

- Maximize between class variance
  - $\max \quad w^T \Sigma_{\text{within}} w$
  - $\Sigma_{\text{within}} = \frac{\sum_{i=1}^{N} \epsilon_i {\epsilon_i}^T}{N-1}$

- the project matrix is the Eigen vectors of matrix $\text{inv}(\Sigma_{\text{within}})\Sigma_{\text{between}}$



Fisher's linear discriminant(2)

- Matlab Code

```
% Compute total covariance matrix
St = cov(X);
% Sum over classes
for i=1:nc
     cur_X = X(labels == i,:);  % Get all instances with class i
    % Update within-class scatter
    C = cov(cur_X);
    p = size(cur_X, 1) / (length(labels) - 1);
    Sw = Sw + (p * C);
end
% Compute between class scatter
Sb = St - Sw;
```

$$\Sigma_{\text{totoal}} = \Sigma_{\text{between}} + \Sigma_{\text{within}}$$

- The project matrix formed by the M Eigen vectors of inv(Sw)*Sb

More stories

- Kernelizing
  - Fisher discriminant analysis with kernels
- Clustering
  - Adaptive Dimension Reduction Using Discriminant Analysis and K-means
- Other extension:
  - Geometric Mean for Subspace Selection
    - LDA: arithmetic mean of kl divergence
    - Replace arithmetic mean with geometric mean
  - Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis
    - Gaussian hypothesis is not hold

Thanks for joining the journey of statistical learning