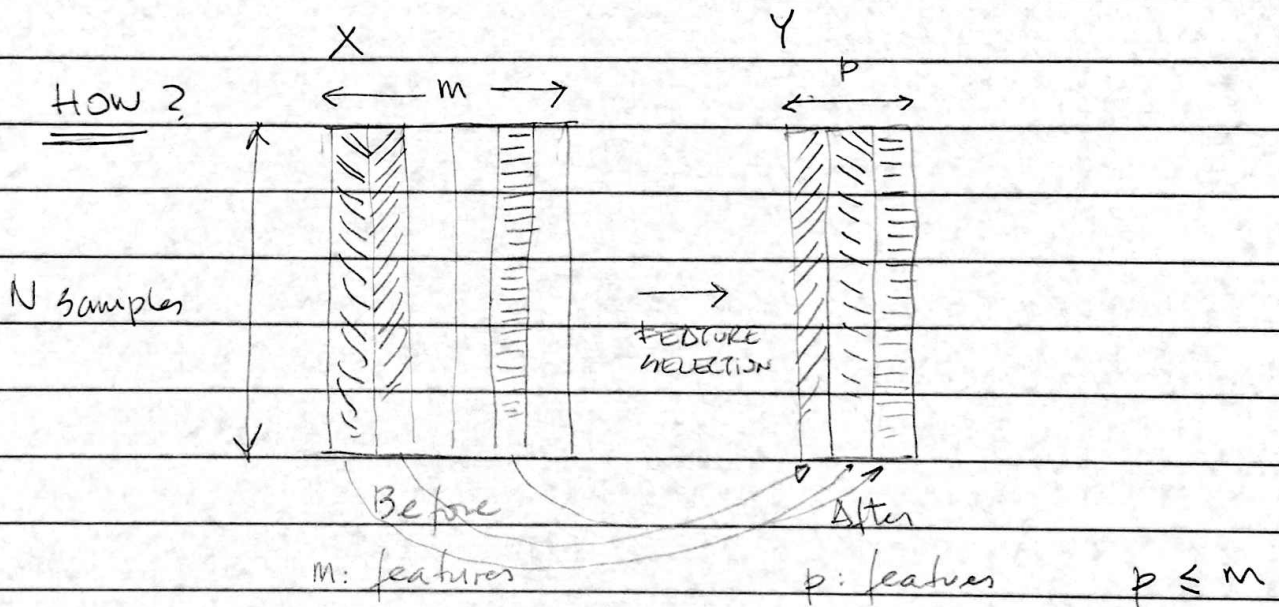


FEATURE SELECTION

① WHY?

- Many extracted features are not "discriminative"
⇒ NO INFORMATION FOR SEPARATION
- Some features are correlated
⇒ REDUNDANT
- The less the better
⇒ CLASSIFICATION WILL BE SIMPLER

② HOW?



Given X find Y
 \downarrow \downarrow
 $N \times m$ $N \times p$

$$Y_j = X_{ij}$$

X_i : i th column of X

Y_j : j th column of Y

MATLAB

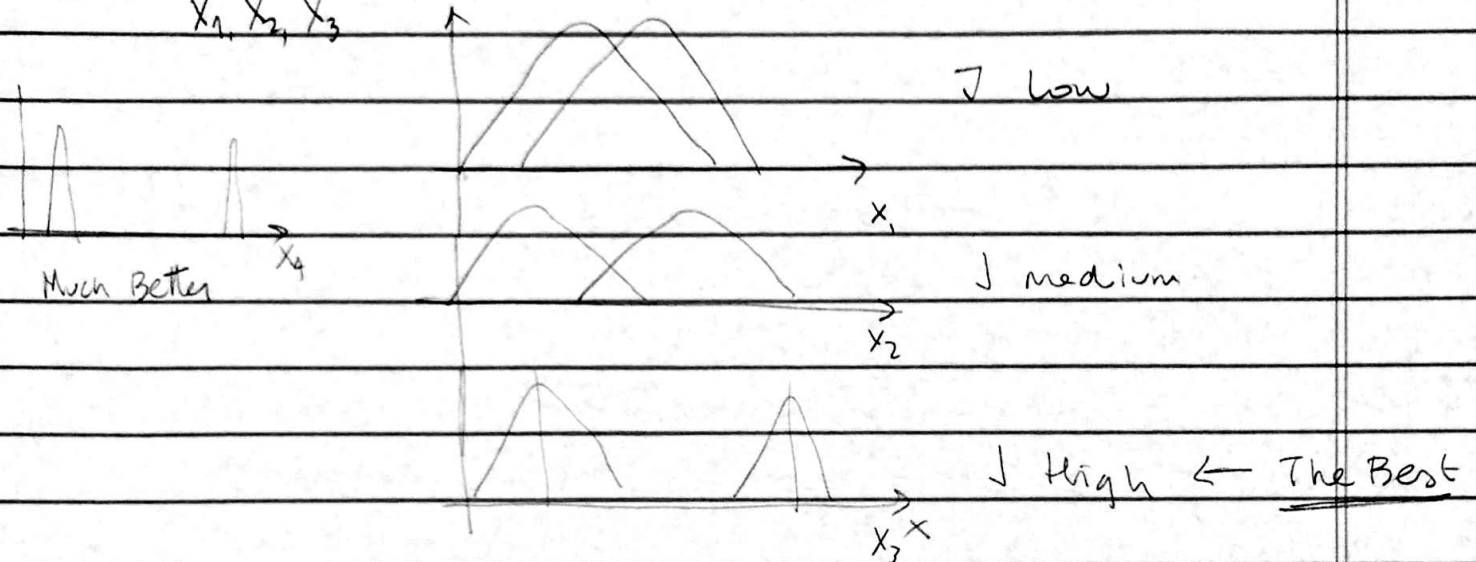
$$Y = X(:, ii);$$

$$ii = [i_1 \ i_2 \ \dots \ i_p];$$

We need a "score", that is an objective function to maximize. We call this function J

Example for one feature: We have three features

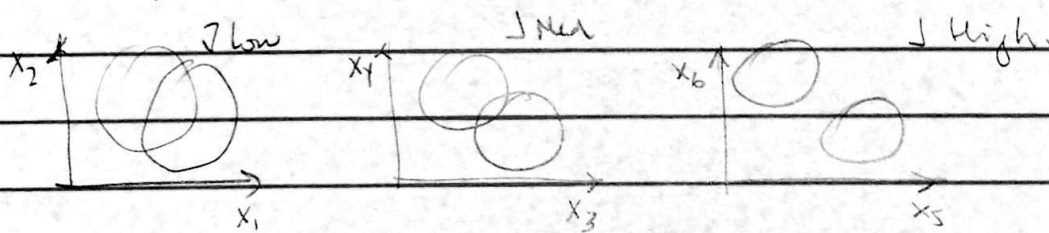
x_1, x_2, x_3



A simple score is Fisher function

$$J(x) = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \rightarrow \text{Max}$$

More features? How could be J for two features



Within-class Covariance

$C_W = \sum$ Dispersion of each class

Between-class covariance

$C_b = \sum$ Dispersion related to the center

→ PAT03_GUI_Fisher.m

$J = \text{trace}(C_W^{-1} C_b)$ → This can be used for more features!

trace: sum of the diagonal

Feature Selection Methods

① Exhaustive Search (Brute-force)

Find the p features of X that maximize J
evaluating all possible candidates

ex: 5 features $\dots \Rightarrow$ 3 features

$$\left. \begin{array}{l} J(x_1, x_2, x_3) \\ J(x_1, x_2, x_4) \\ J(x_1, x_2, x_5) \\ \vdots \\ J(x_3, x_4, x_5) \end{array} \right\} \rightarrow \text{Max}$$

② SFS

SFS-Land } 4th

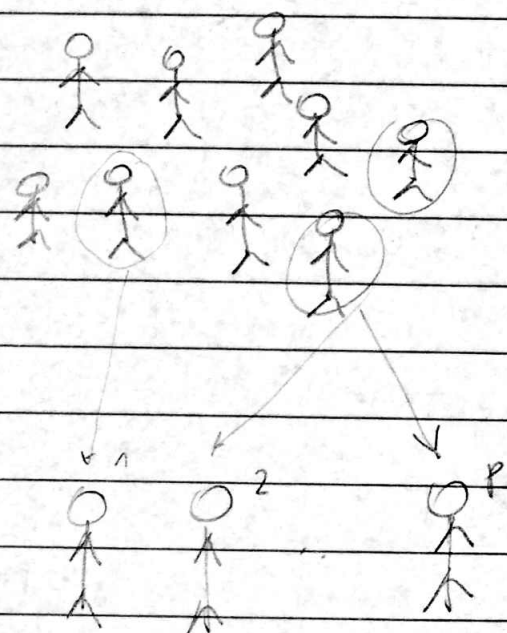
SFS-Law

① You can get married
 p times (one by one)

② You cannot get divorced

③ Try to build the

(TEAM!) "best" marriage with p "spouses".



(IF SPOUSE 1 IS GOOD ENOUGH FOR
JOB X, SELECT SPOUSE 2 FOR ANOTHER JOB)

SFS - Algorithm

Sequential Forward Selection

① Select the feature i_1 of X so that

$$J(X_{i_1}) \rightarrow \text{Max} \quad \text{for } i_1 = 1 \dots m$$

② Select the feature i_2 of X so that

$$J(X_{i_1}, X_{i_2}) \rightarrow \text{Max} \quad \text{for } i_2 = 1 \dots m \\ i_2 \neq i_1$$

③ Select the feature i_3 of X so that

$$J(X_{i_1}, X_{i_2}, X_{i_3}) \rightarrow \text{Max} \quad \text{for } i_3 = 1 \dots m \\ i_3 \neq i_1, i_3 \neq i_2$$

⋮

until p -step

Sequential

③

SBS - Backward Selection

Law:

- ① You are married with everyone
- ② You get divorced $m-p$ times
- ③ You cannot get married again
- ④ Try to build a team

④

Plus & Minus

- SFS l times
- SBS r times
- repeat until p features are selected

$l > r$

⑤

Branch & Bound

$S = [0 \ 0 \ 0 \ 0 \ 0 \ \dots \ 0];$

selected features

m

$J_{\max} = 0; \ Y = [];$

for $k = 1 : p$

SFS

for $j = 1 \dots m$

if $S(j) = 0$

$J = \text{fisher}(Y, X_j)$

if $J > J_{\max}$

$j_s = j$

$J_{\max} = J$

end

end

$Y = [Y \ X_{j_s}]$

end end $S(j_s) = 1$