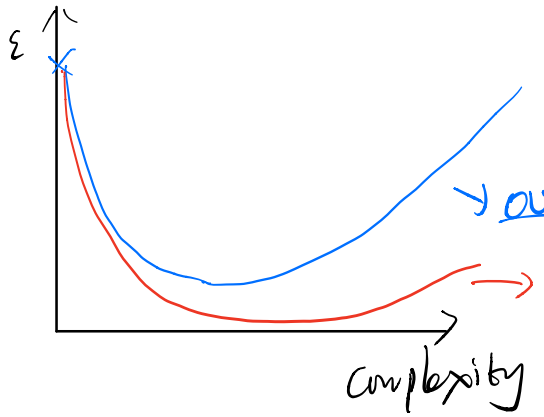
Feature selection

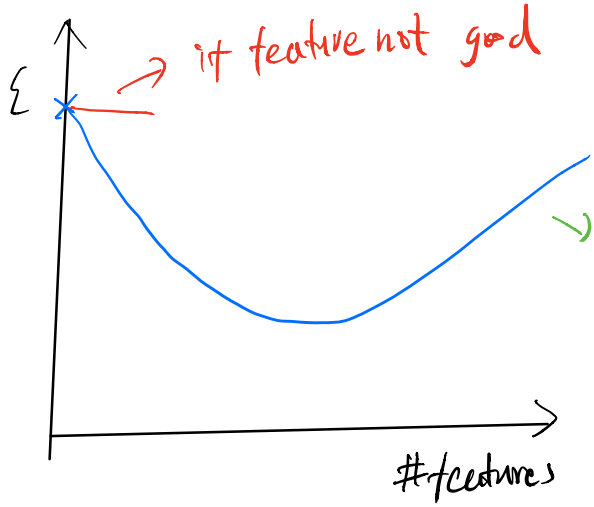
1.



↘ overfitting

→ if we have more observations

2.



→ if feature not good

↘ also depends on the order of feature (importance of feature)

3. L1 → LASSO → $\|w\|$ → avoid overfitting complexity

↓ reduce features ◇ ☆

why ?

↘ random selection

$\lambda \|w\|$

4.

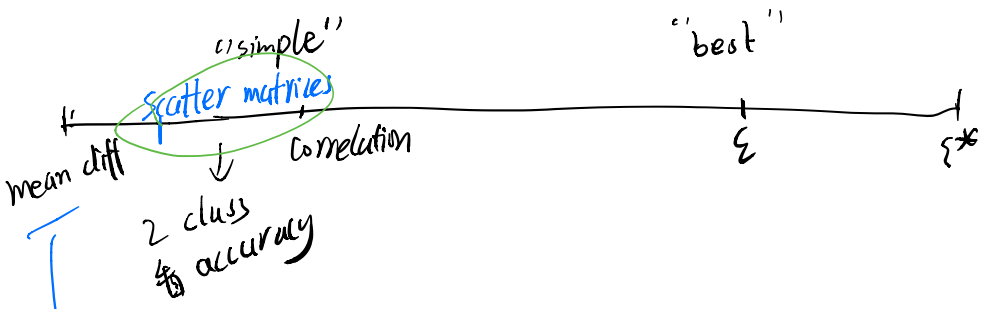
100 features → how to choose

① → fit classifier → check error
wrapper
evaluate.
cross validation

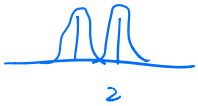
② Bayes error ? → not go down, #features↑

↓

selection criteria

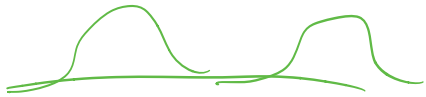


doesn't work



tends to 2

mean μ



$(m_1 - m_2)^2$ → criterion

σ^2

ficial criterial?

scale data

M_1 中, σ (m_1 m_1)

compare two features

of that function

which is better

5. scatter matrices (covariance matrices)

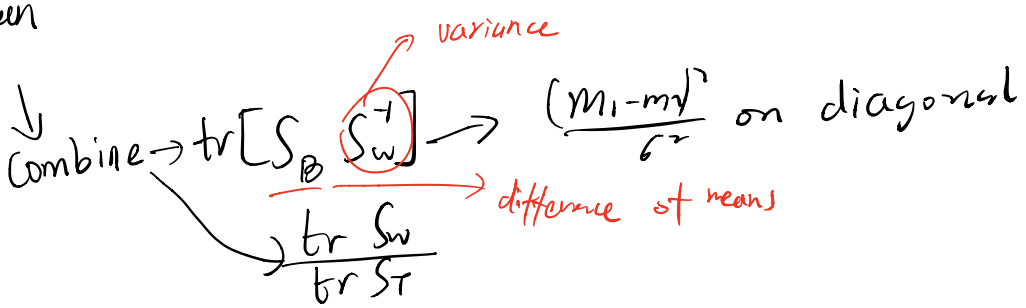
avg covariance

S_{within} take by class

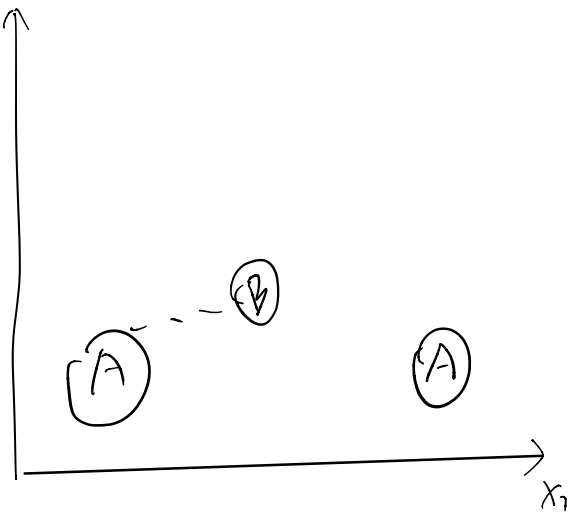
S_{Total} take all data

S_{Bias} only takes mean of class?

Between



6. X_2

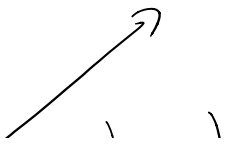


facial criterial

choose X_2

KNN

7. If criterion fired, how to find my best feature subset.



10 Dimentional → search strategy, change criterion

$\binom{20000}{10} \approx (2^{0000})^{10}$

8. Individual FS → (not correlated) ← feature selection

9. forward selection

{1, 2, 3, 4, 5}

1st → choose 4 (best feature) → {4, 1) (4, 2) (4 3)

include another and also check correlation

(4, 5) → see correlation

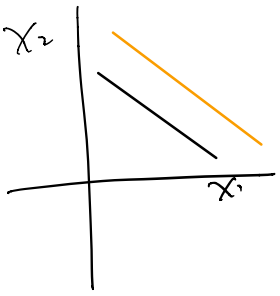
2nd → (4, 3, 1) (4, 3, 2) (4, 3, 5)

check correlation find 3rd best

10. backward selection:

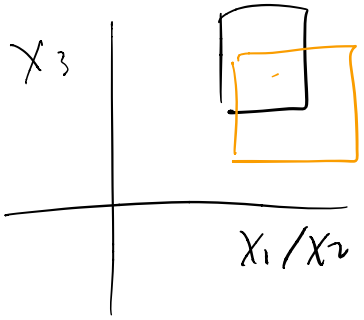
start from whole features → remove features

problem of 9.

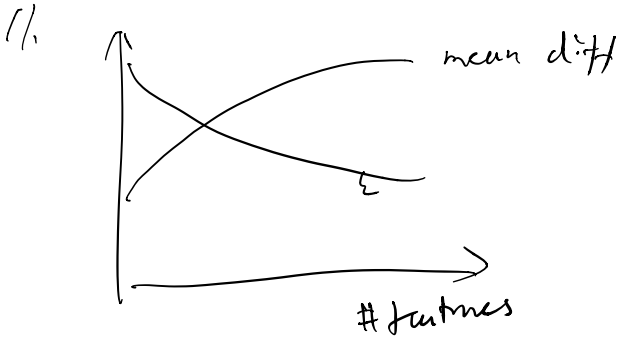


?

choose X3 first, then choose 1 or 2



↓ only choose 2.


mean diff
ε
Iterations