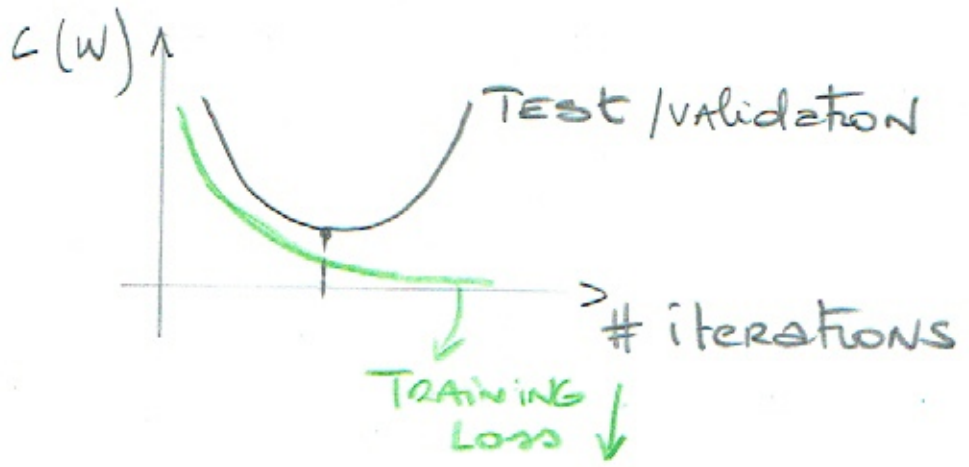
REGULARIZATION

↳ prevent overfitting by reducing complexity of the model



STRATEGIES ① EARLY STOPPING !

② Penalize model for complexity during training.

$$C(w,b) = \frac{1}{m}\sum (y_p - y)^2 + \text{Complexity measure}$$

↓

give preference to 'smaller' weights → AROUND ZERO

L_2 - regularization (Ridge regul.) → square of l_2-norm $\vec{w}$

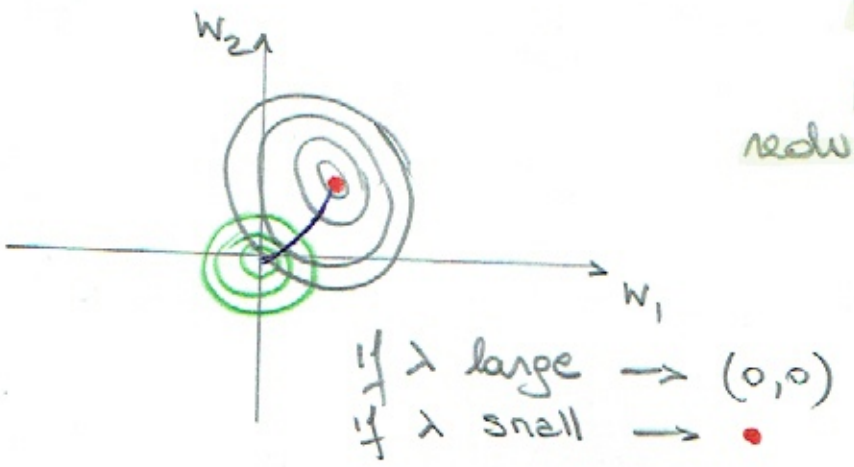
$$C(w,b) = \frac{1}{m}\sum (y_p - y)^2 + \lambda \sum w_i^2$$

↓

hyperparameter !

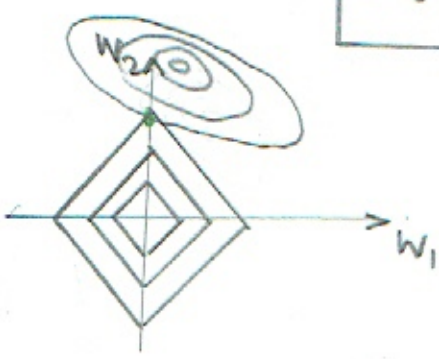
↓

balancing between reducing TRAINING ERROR vs. model complexity.



L_1 - regularization (Lasso reg)

$$C(w,b) = \frac{1}{m} \sum (y_p - y)^2 + \lambda \sum \|\bar{w}\|_1$$



krank w_2 up AND put $w_1 = 0$
→ very deterministic
→ sparse vector!

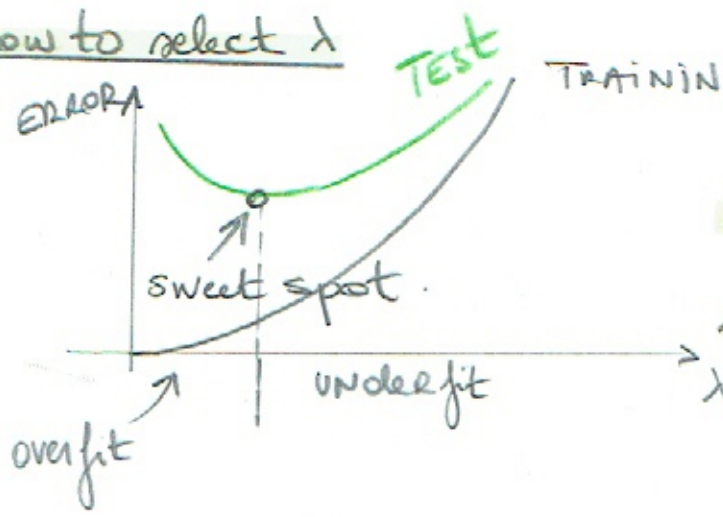
decrease λ and see what feature pops up 1st, 2nd, ...

↳ when enough knowledge → L_2

Elastic Net ←

$$\lambda |w|_1 + \mu \|w\|_2$$

ERROR + how to select λ



$\lambda = 0$ → TRAINING ERROR ≈ 0

$\lambda >>$ → λ dominates
→ simple model
→ TRAINING ERROR high

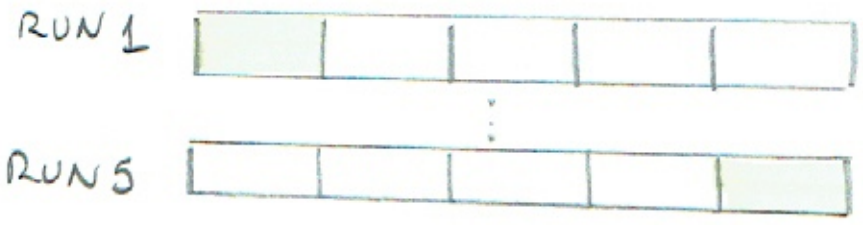
$\lambda <<$ → focus on training
→ test ERROR high!

TRAINING	CROSS VAL	Test

↳ evaluate which λ gives you lowest cross VAL ERROR

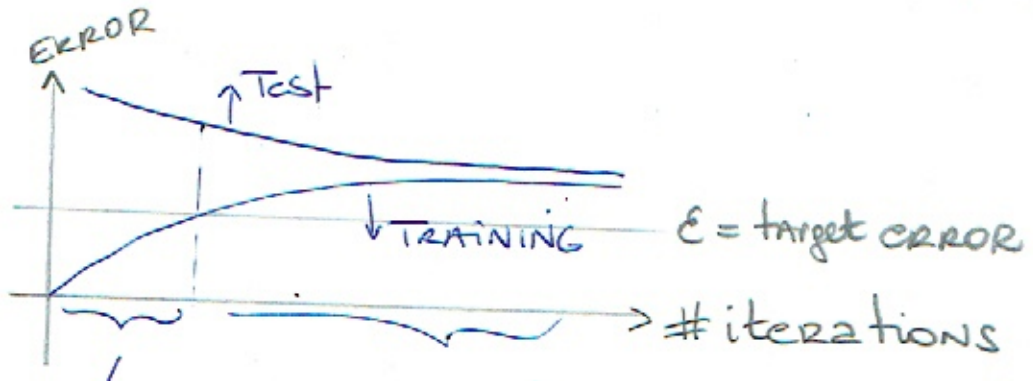
⚠ If too many actions on cross VAL set
→ Risk of overfitting!

→ k-fold cross validation



k-validation ERRORS
↓
common to use
CV = 5

Debugging



High Variance

* TRAINING ERROR < ε
* test ERROR too high

SHRINK GAP

1. add more data
2. reduce model complexity
3. INCREASE λ
4. BAGGING

High BIAS

* TRAINING ERROR > ε
  → model is wrong
  → Adding data is useless
* Test ERROR is never lower than TRAINING ERROR

1. Make model more complex
2. Add more features
3. decrease λ
4. Boosting