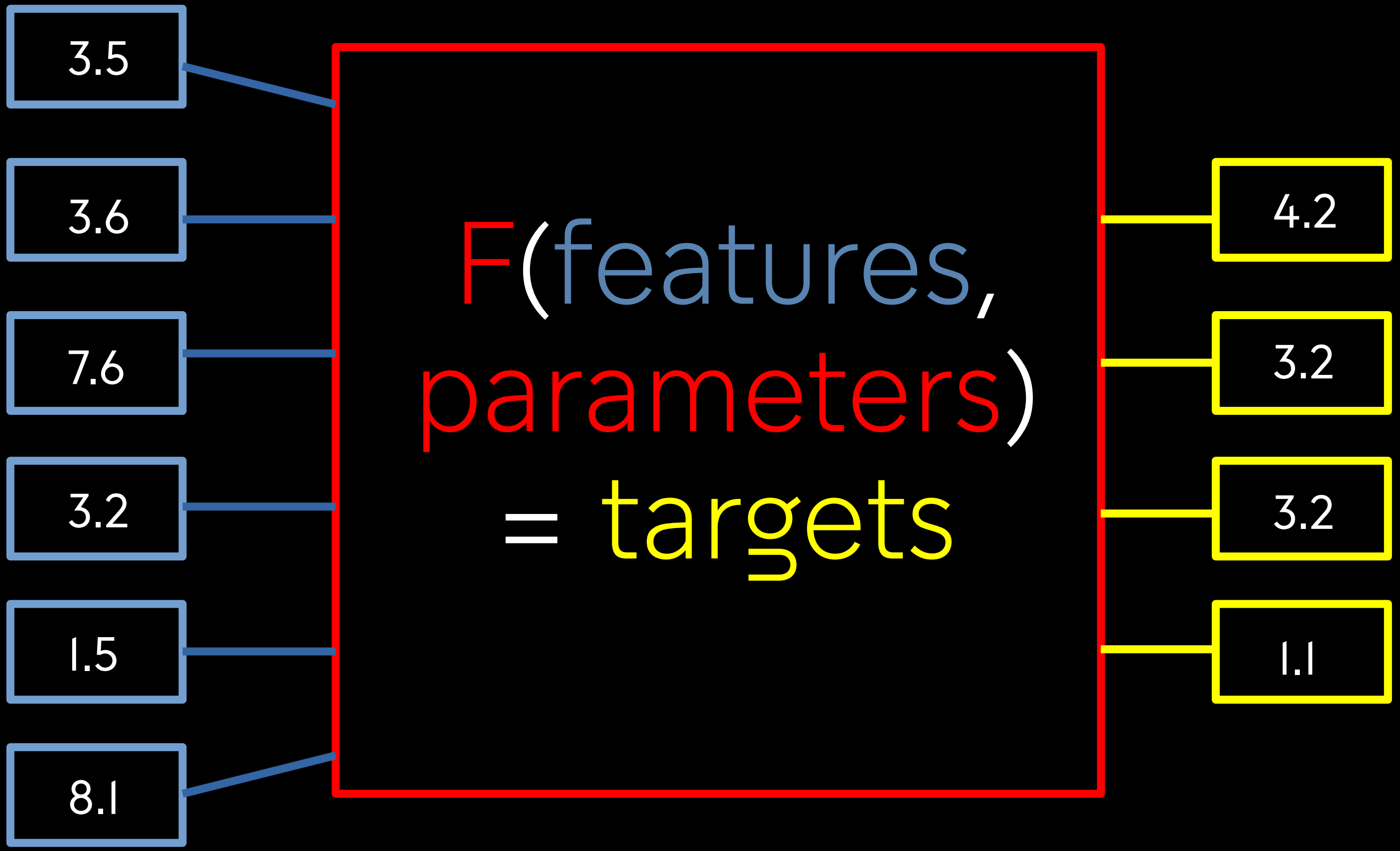
Machine Learning
Regression Problem

Thomas Haschka
5 Mai 2021

Regression model

- Fit input parameters to expected output parameters
- Typical fit function

Regression Model



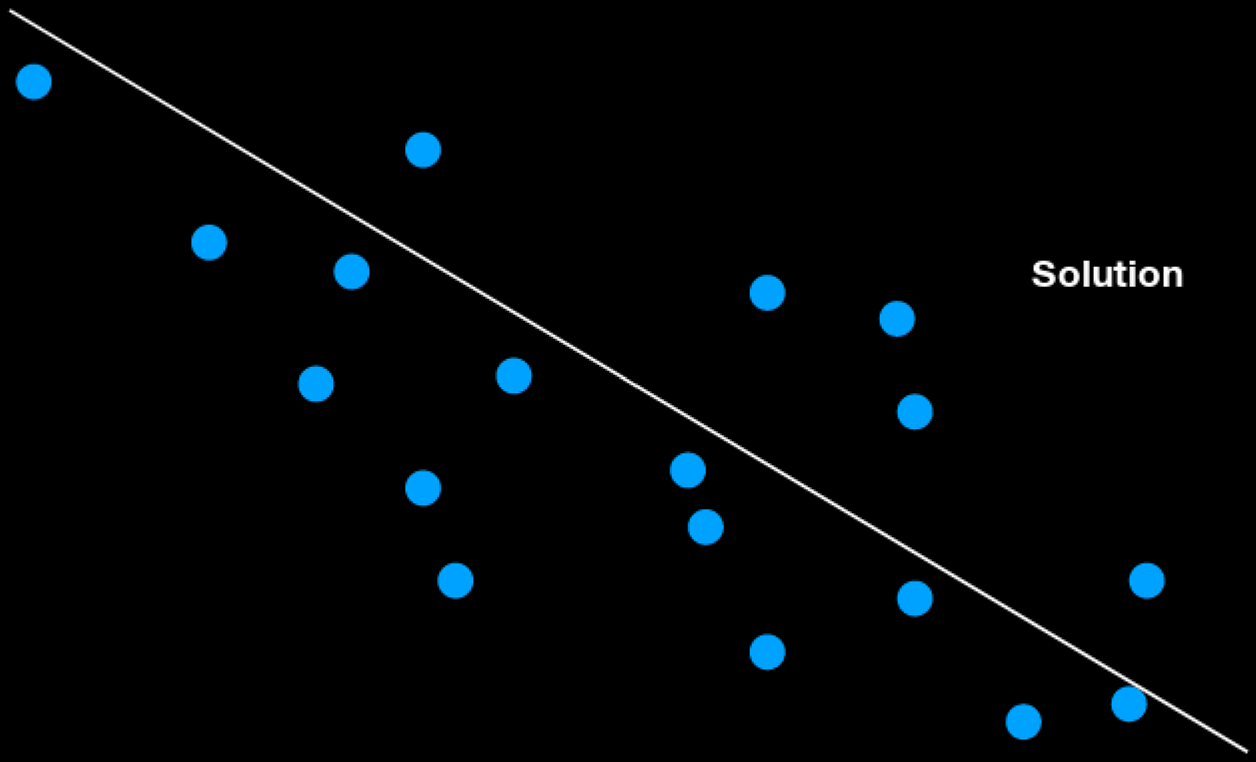
Loss Function

Distance(Estimated,Truth)

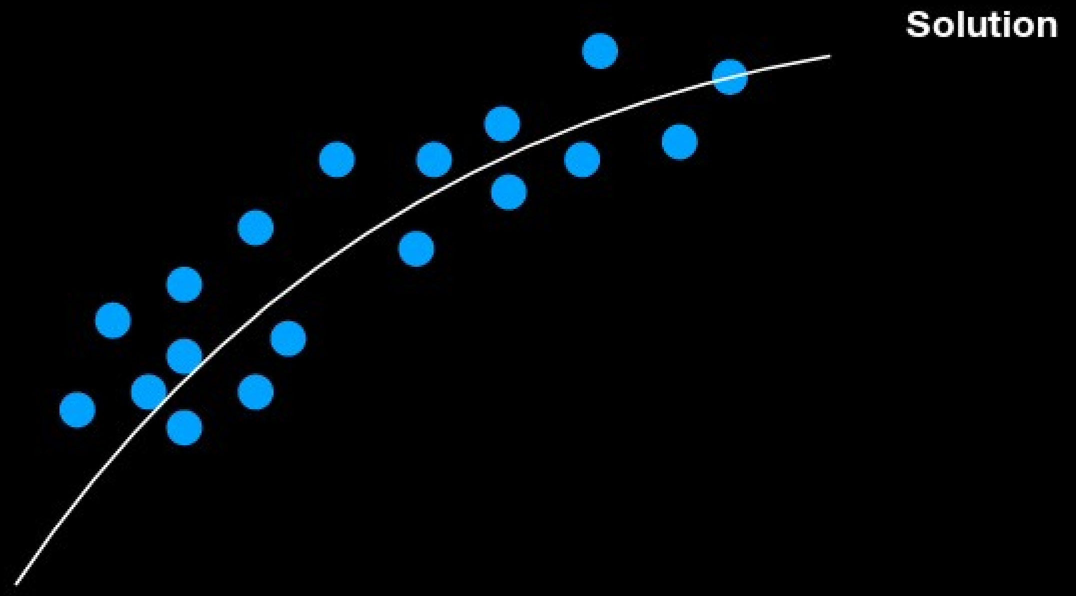
Distance(F(Features,Parameters), Truth)

In adapting F and its parameters
we minimize the distance, between
prediction and truth

Example Linear Regression



Example polynomial fit



Regularisation

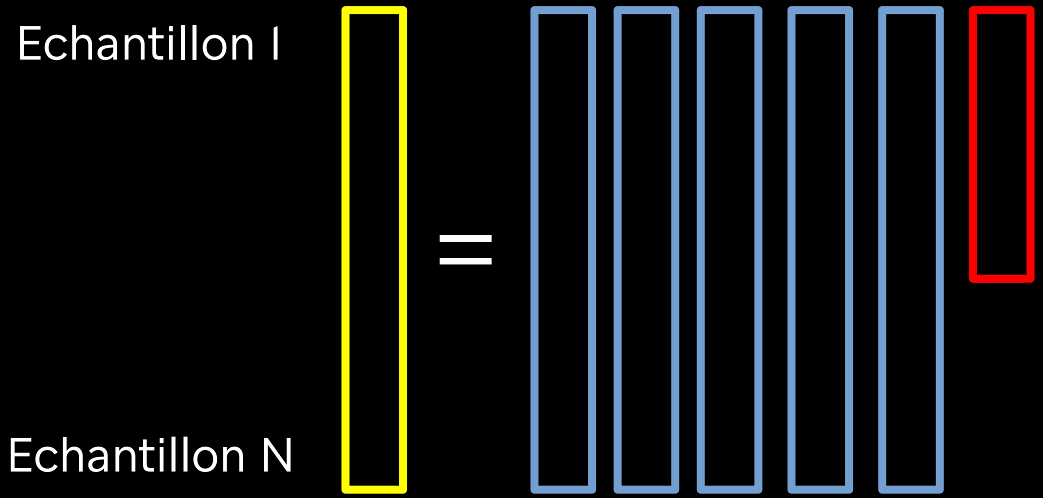
Linear Regression: $Ax=b$

Loss Function: $||Ax-b||$

Regularized Loss Function:

$$||Ax-b||+|\lambda^1 x|+||\lambda^2 x||$$

Regularization



Regularized Loss Function:

$$\Sigma||Ax-b||+|\lambda^1 x|+||\lambda^2 x||$$

L1 Part L2 Part

Regularization

Regularized Loss Function:

$$\Sigma||Ax-b||+|\lambda^1 x|+||\lambda^2 x||$$

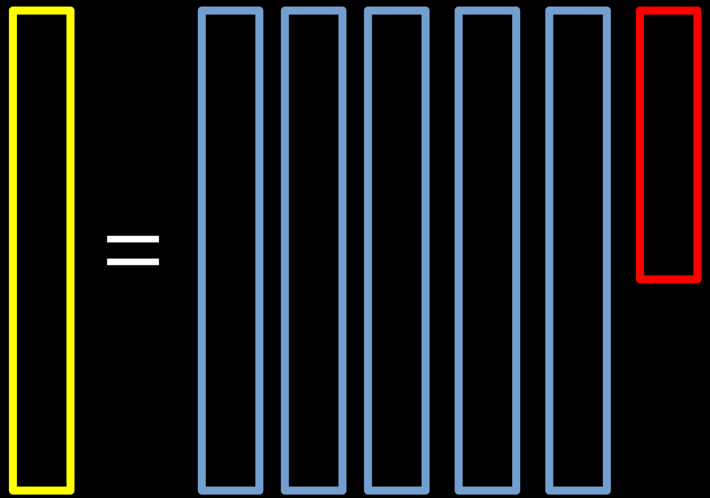
L1 Part

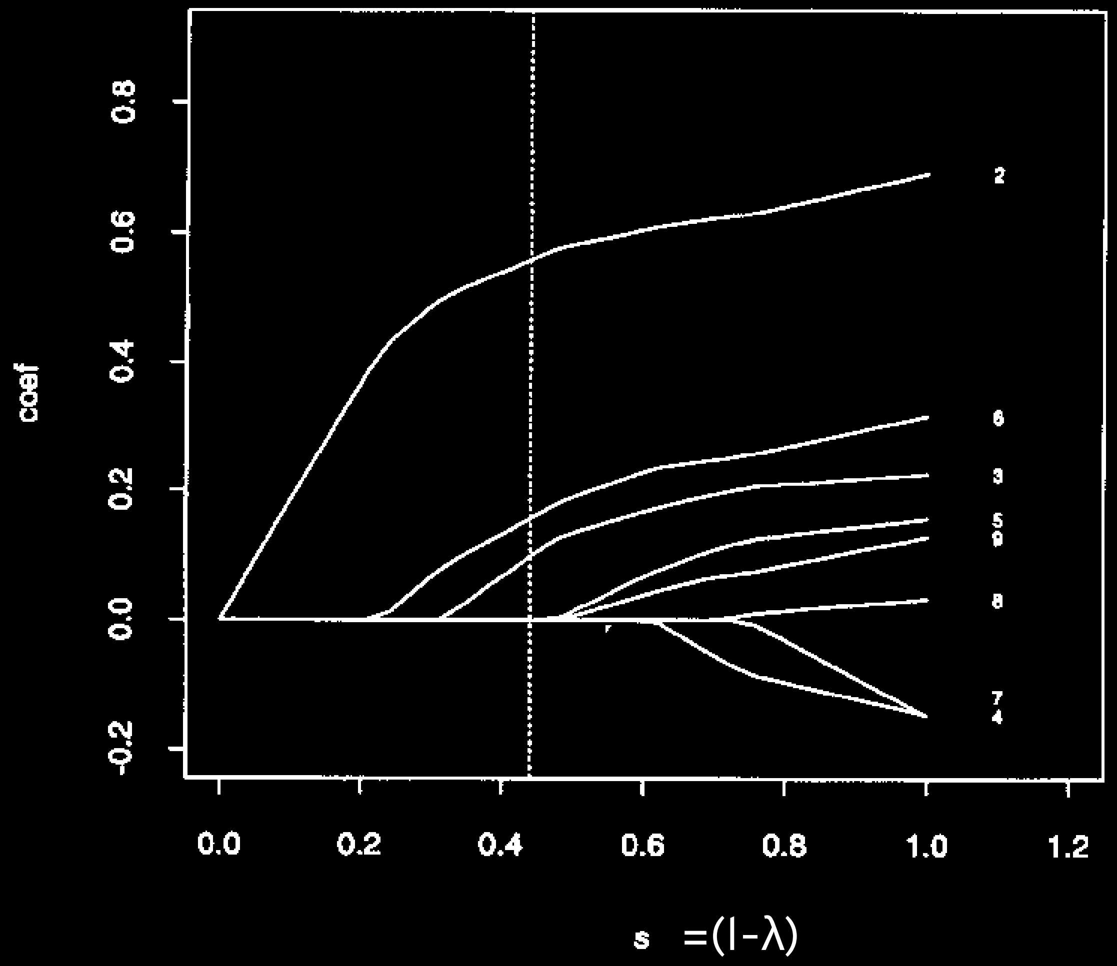
makes solutions sparse

=>fewer features matter 0s in X

L2 Part

keeps solutions small



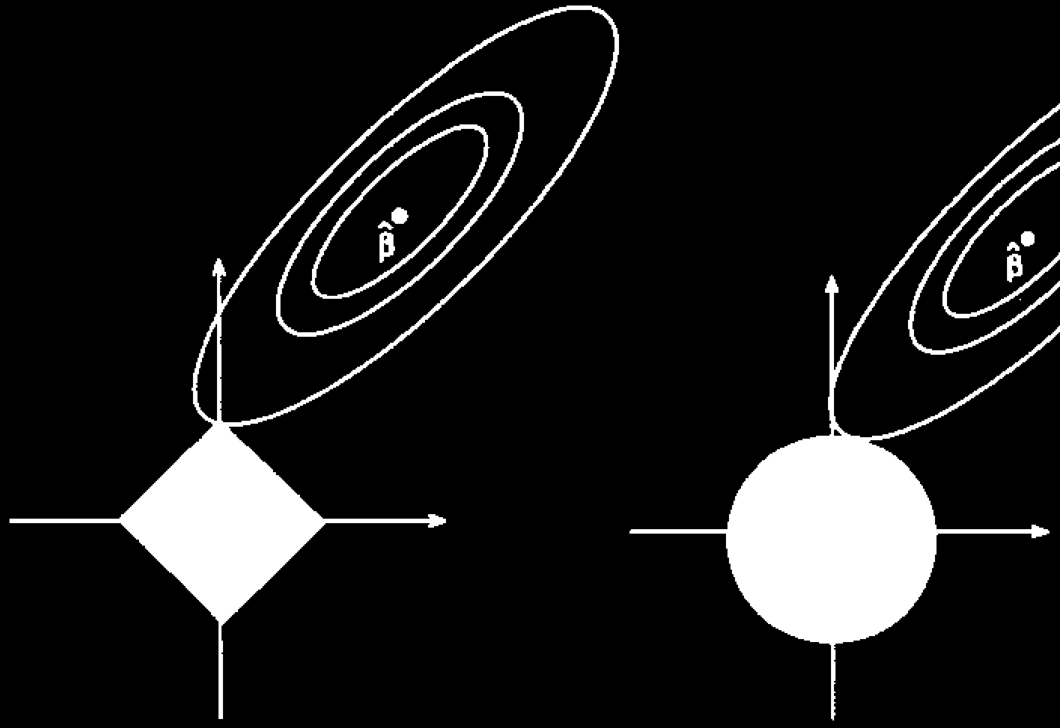


Evolution of values
in x by modifying λ

R. Tibshirani et al. (1996)

Lasso:

$$\Sigma||Ax-b||+|\lambda' x|$$



(a) (b)

R. Tibshirani et al. (1996)

Lasso:

$$\Sigma||Ax-b||+|\lambda' x|$$

Stability against outliers

- L1-norm (Absolute Value)

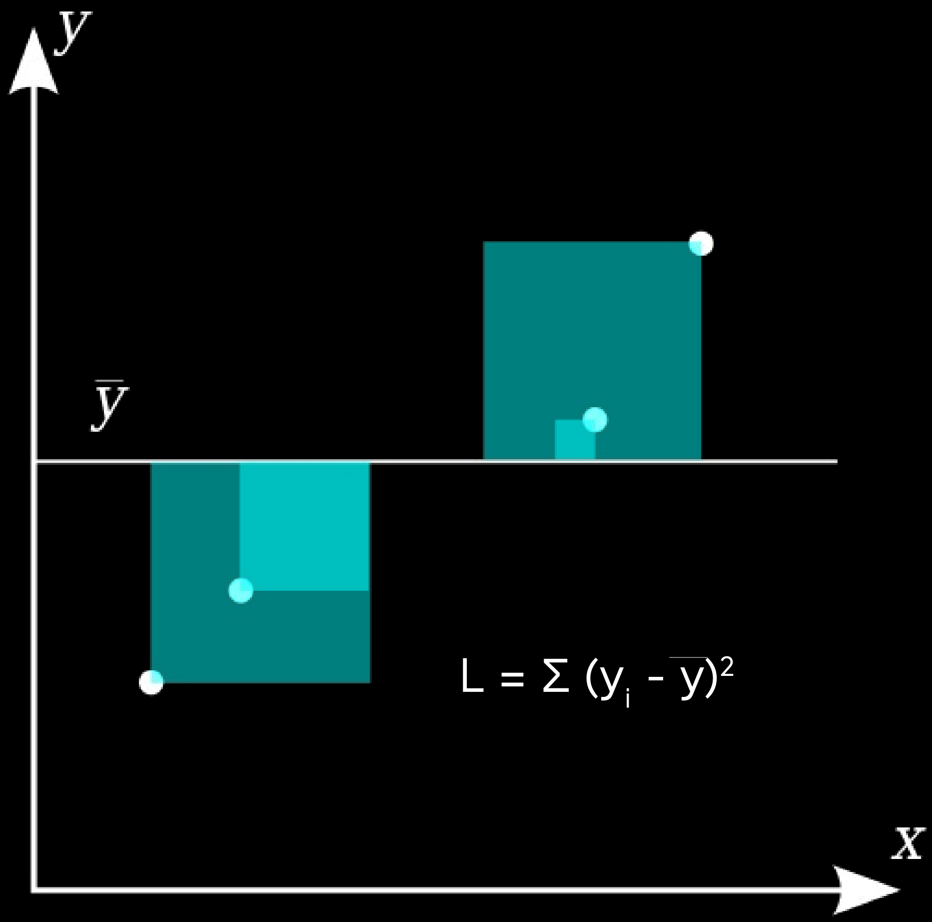
  less sensitive to outliers (can yield better fits)

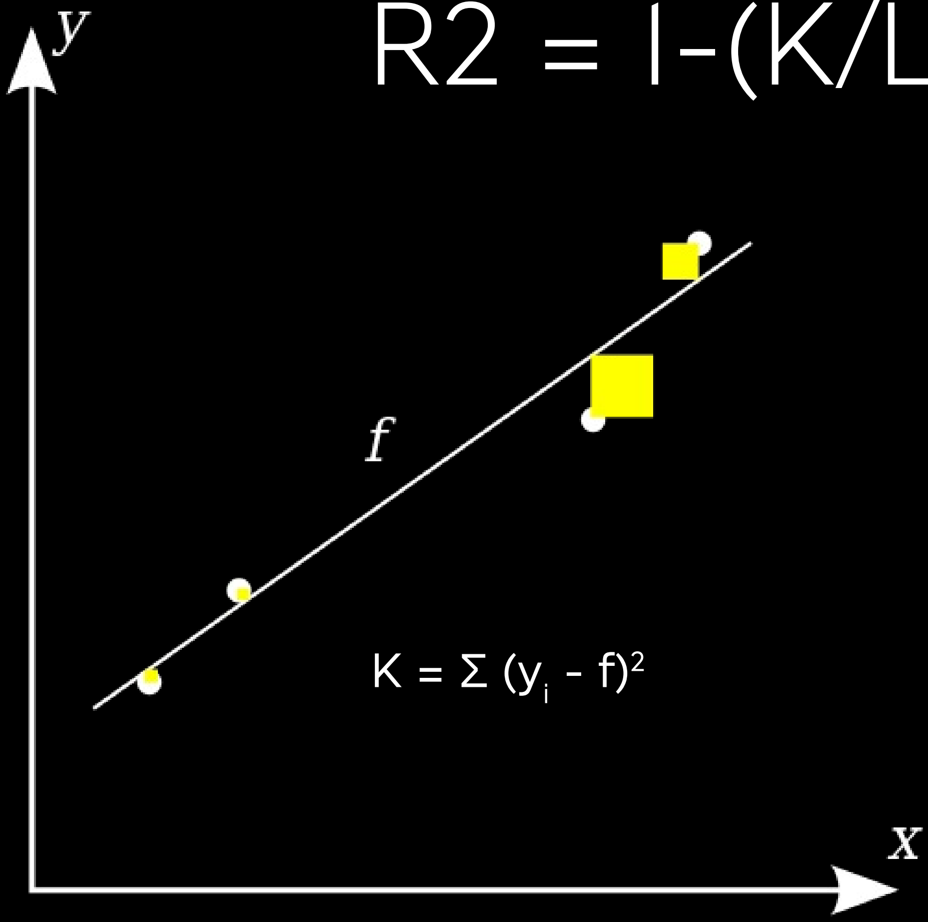
- L2-norm (Euclidean Norm)

  sensitive to single large outliers

Coefficient of Determination

R2 = l-(K/L)





Model Validation

Dataset

Trainingset | Testset

1 | 2 | 3 | 4

n-fold cross validation

Train On

Test On	1	2	3	4
1	0.9	0.7	0.5	0.9
2	0.5	0.8	0.6	0.6
3	0.6	0.6	0.9	0.7
4	0.7	0.2	0.5	0.9

yields a confidence SCORE +- error
->How often will I hit the right target

Final Model Validation
only with Testset

if possible cross validated

cross validation datasets
can overlap
and be chosen randomly

but know what you are doing!