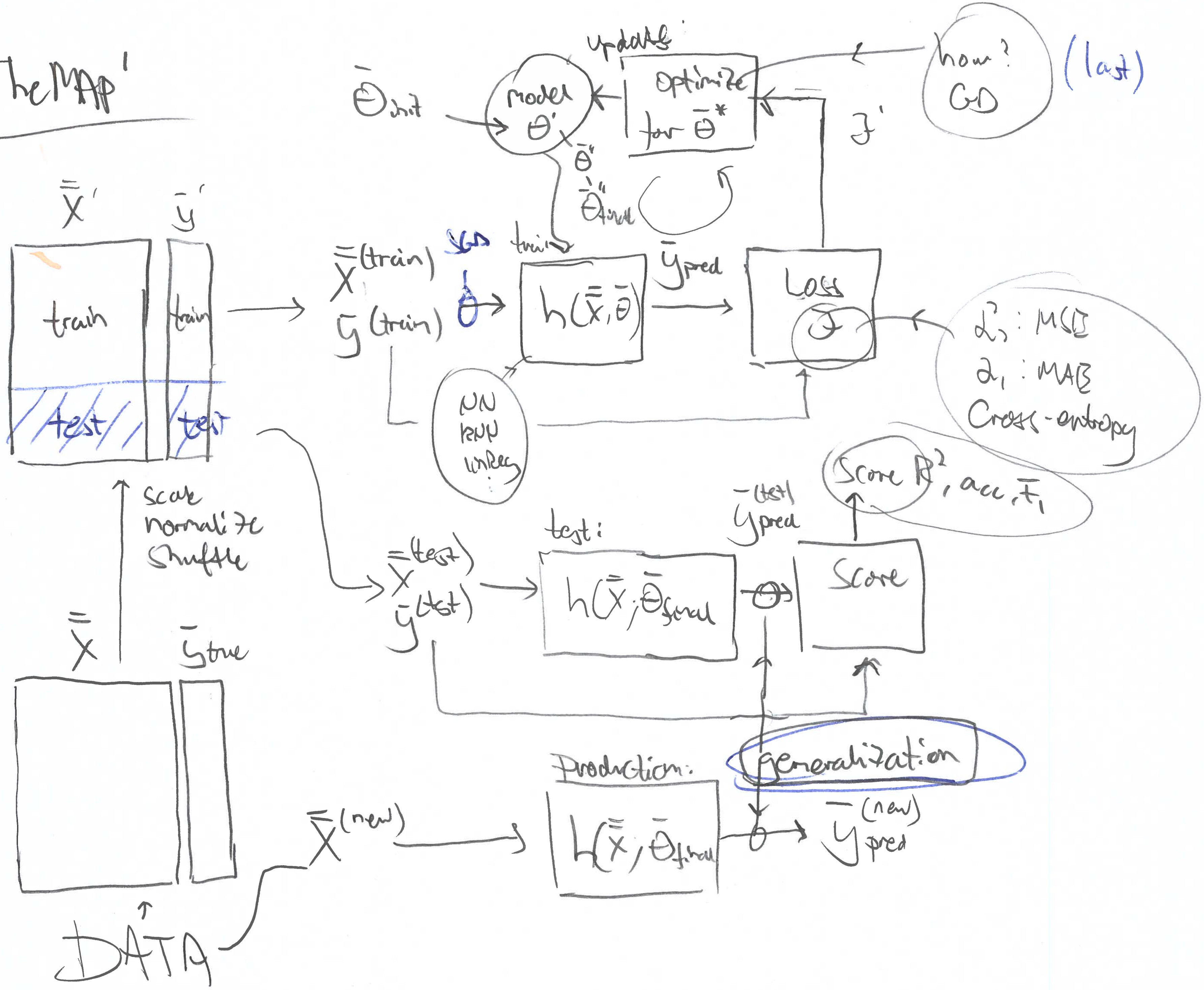
ITMAL L03

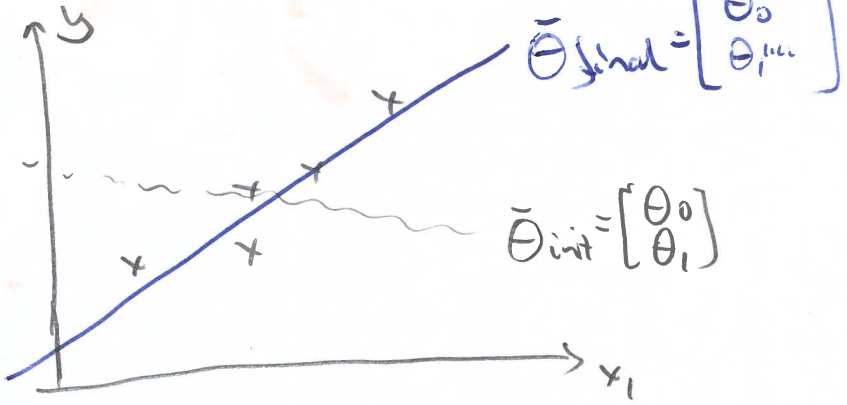
Notes

'The MAP'

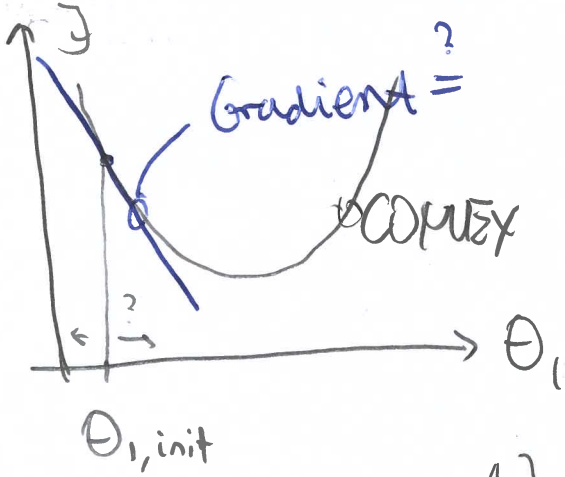

train
test
train
test
scale
normalize
shuffle
y true
DATA
X (train)
y (train)
SGD
train
Θ init
model
updating:
optimize
for Θ*
Θ final
h(X, Θ)
y pred
Loss
J
how?
GD
(last)
NN
KNN
linReg
MSE
MAE
Cross-entropy
Score R², acc, F1
test:
X (test)
y (test)
h(X, Θ final)
y pred (test)
Score
Production:
X (new)
h(X, Θ final)
y pred (new)
generalization

Gradient Descent

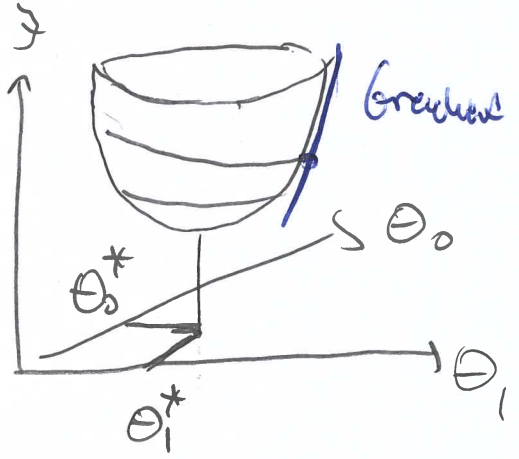
DATA-Space, 1D, lin. reg



PARAMETER-Space, 1D



P-Space, 2D



1D: $\theta_1^{(next)} = \theta_1 - \eta \frac{\Delta J}{\Delta \theta_1}$

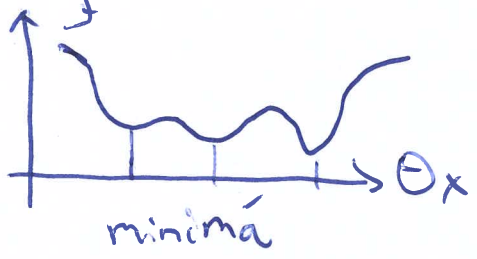
N-Dim gradient $\stackrel{?}{=} \nabla_{\bar{\theta}} J = \begin{bmatrix} \frac{\partial}{\partial \theta_0} \\ \frac{\partial}{\partial \theta_1} \\ \vdots \end{bmatrix} J$

ND: GD:

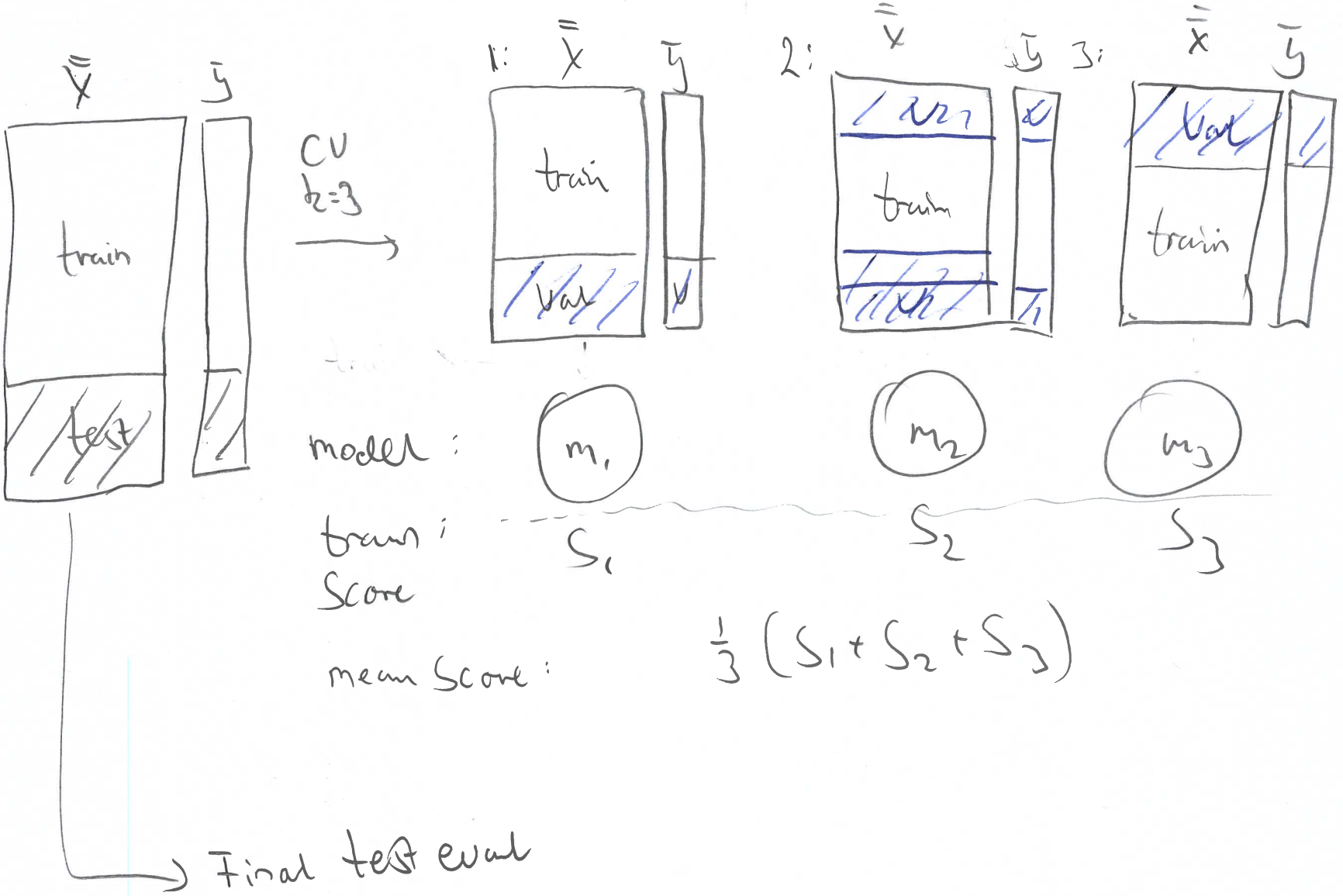
$$\bar{\theta}^{(next)} = \bar{\theta} - \eta \nabla_{\bar{\theta}} J$$

expensive? + SGD

Non-Convex:



Model Selection via K-fold CV



Score, R^2, coeff. of determination

Range for R^2 ?

$$R^2 = 1 - \frac{u}{v}$$

$$u = \sum (\bar{y}_{true} - \bar{y}_{pred})^2 = \underbrace{\| \bar{y}_{true} - \bar{y}_{pred} \|_2^2}_{L_2^2}$$

$$v = \sum (\bar{y}_{true} - \bar{\mu}_{true})^2 = \| \bar{y}_{true} - \bar{\mu}_{true} \|_2^2$$

↑ scalar to vector

u_{max} for $\bar{y}_{pred} != \bar{y}_{true} \Rightarrow \infty$

u_{min} for $\bar{y}_{pred} == \bar{y}_{true} \Rightarrow \emptyset$

$v_{max} = ?$

$v_{min} = ?$

$$R^2|_{u_{max}} = 1 - \frac{\infty}{v} \simeq -\infty \quad \text{worst ?}$$

$$R^2|_{u_{min}} = 1 - \emptyset = 1 \quad \text{best ?}$$