(Single-label) Classification, ANN

Sunday, September 25, 2016 10:50 PM

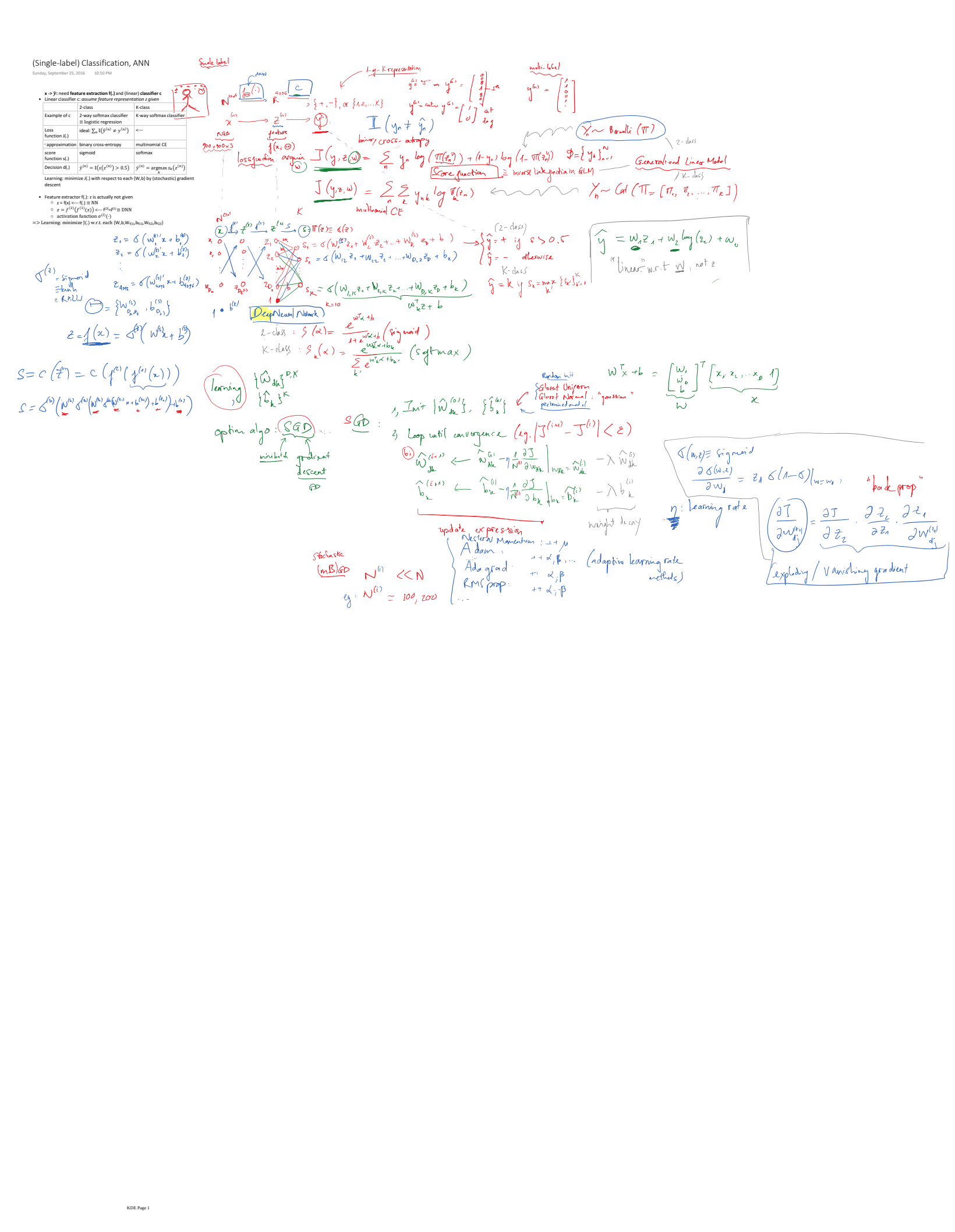
- **x -> ŷ:** need **feature extraction f(.)** and (linear) **classifier c**
- Linear classifier c: *assume feature representation z given*

	2-class	K-class
Example of c	2-way softmax classifier ≡ logistic regression	K-way softmax classifier
Loss function J(.)	ideal: $\sum_n \mathbb{I}(\hat{y}^{(n)} \neq y^{(n)})$	<---
- approximation	binary cross-entropy	multinomial CE
score function s(.)	sigmoid	softmax
Decision d(.)	$\hat{y}^{(n)} = \mathbb{I}(s(z^{(n)}) > 0.5)$	$\hat{y}^{(n)} = \underset{k}{\text{argmax}}\ s_k(z^{(n)})$

Learning: minimize J(.) with respect to each {W,b} by (stochastic) gradient descent

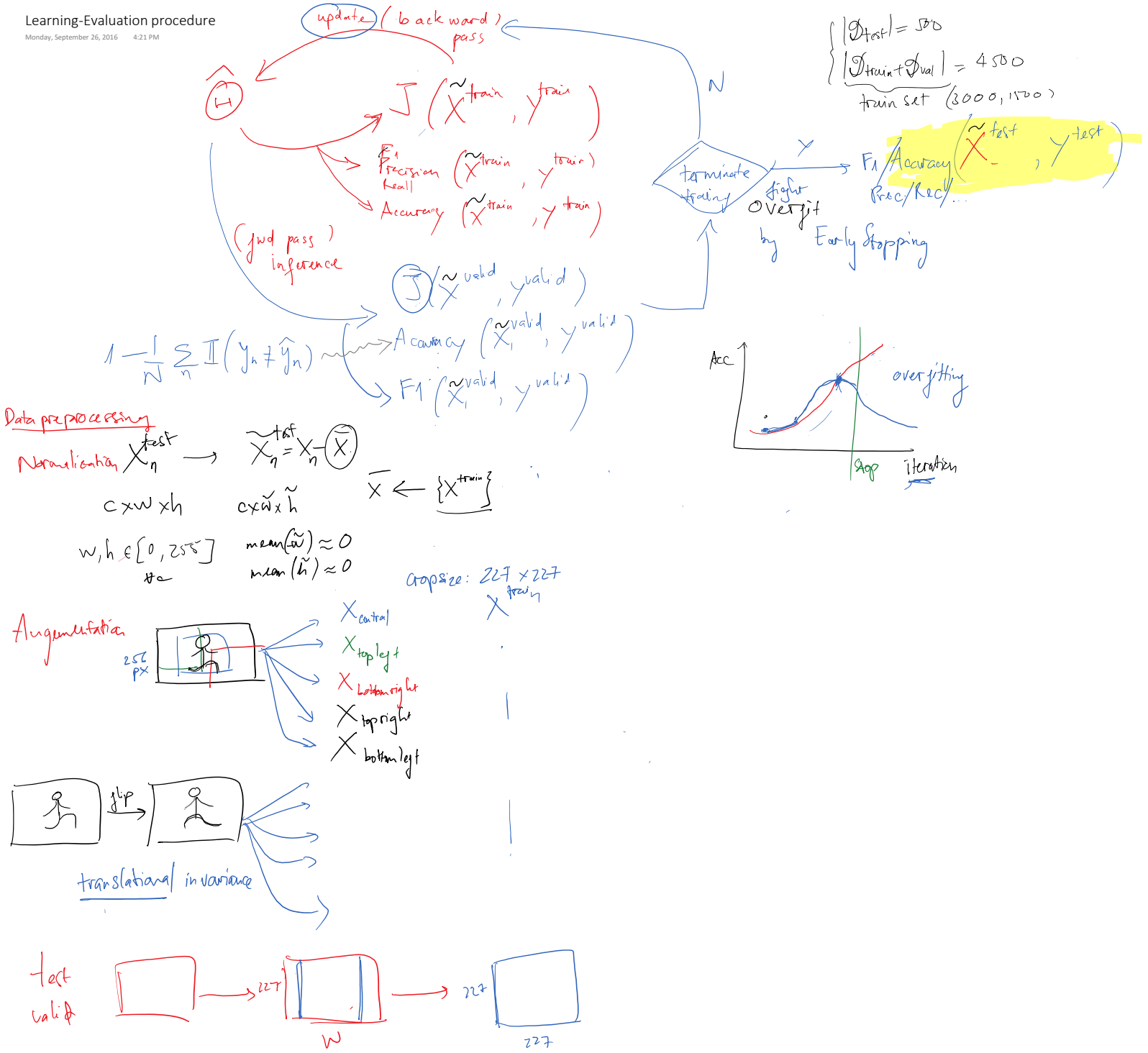
- Feature extractor f(.): z is actually not given
  - $z = f(x) \leftarrow\!- f(.) \equiv$ NN
  - $z = f^{(2)}(f^{(1)}(x)) \leftarrow\!- f^{(2)} \circ f^{(1)} \equiv$ DNN
  - activation function $\sigma^{(l)}(\cdot)$

=> Learning: minimize J(.) w.r.t. each {W,b,W_f(1),b_f(1),W_f(2),b_f(2)}



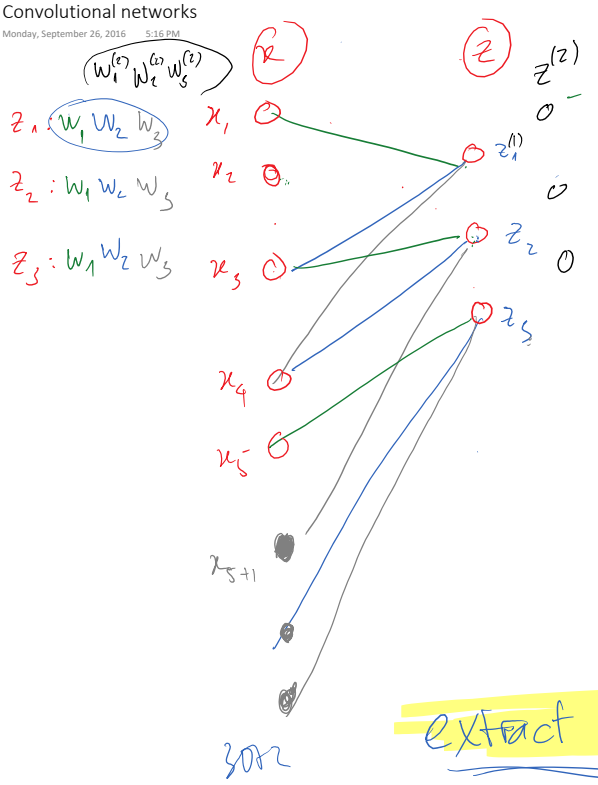
Learning-Evaluation procedure

Monday, September 26, 2016 4:21 PM



Convolutional networks

Monday, September 26, 2016 5:16 PM



Convolutional layer

+ locally connected >< fully-connected (FC)
+ weight sharing
+ hyperparam of the structure
  - stride : 2
  - zero-padding : 3
  - filter size : 3
  - # filters (# kernels) : 2

eg 2 (cs231n)

2 x 2

1

3 x 3 x (3) ← kernel width of the input

2

extract invariant features w.r.t translation

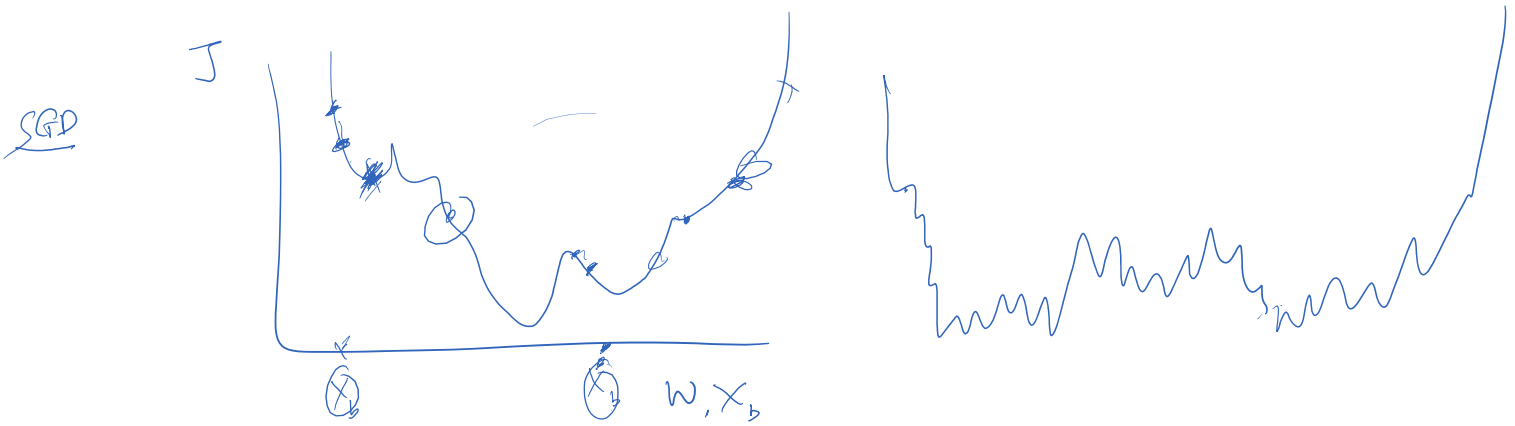
due to { maxpooling / weight sharing

POOL layer

z_{conv} z_{pool}

stochastic optimization

Monday, September 26, 2016 6:33 PM



Stochastic optim

non-convex loss fn

Multi-label learning strategy

Monday, September 26, 2016 6:47 PM

MLMC : OvA