Multi-Label Speaker Recognition using Recurrent Neural Networks

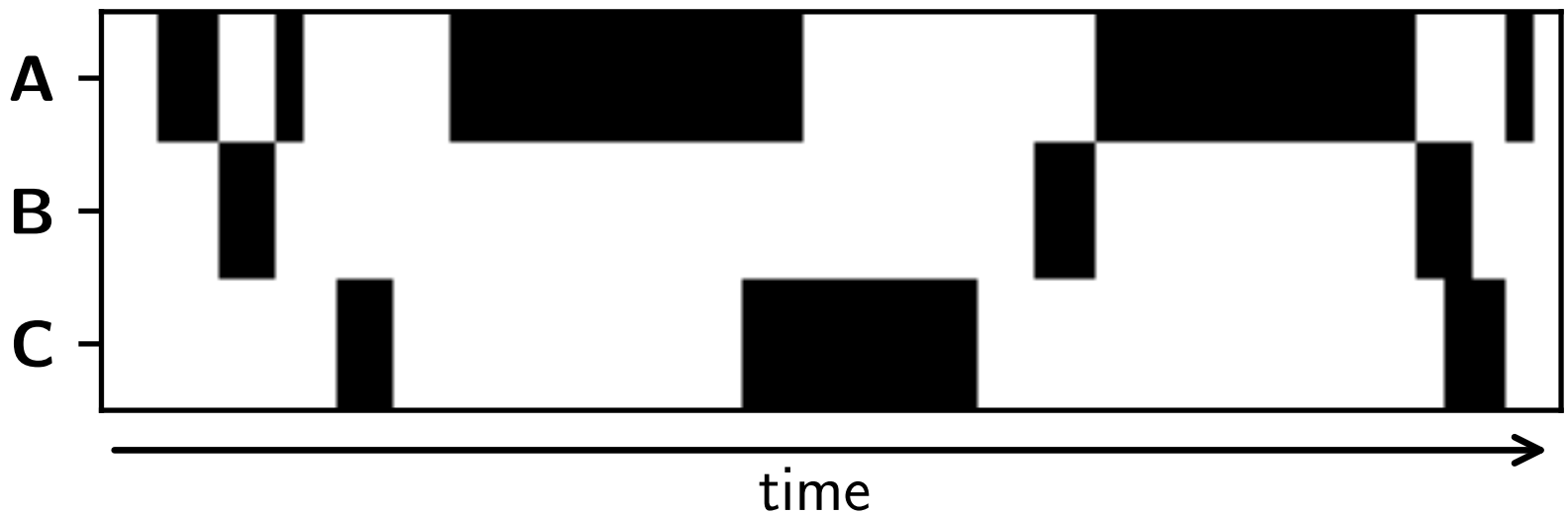
Puhujien tunnistus takaisinkytkeytyvillä neuroverkoilla

Lauri Niskanen

Diplomityöseminaari, 29.10.2018

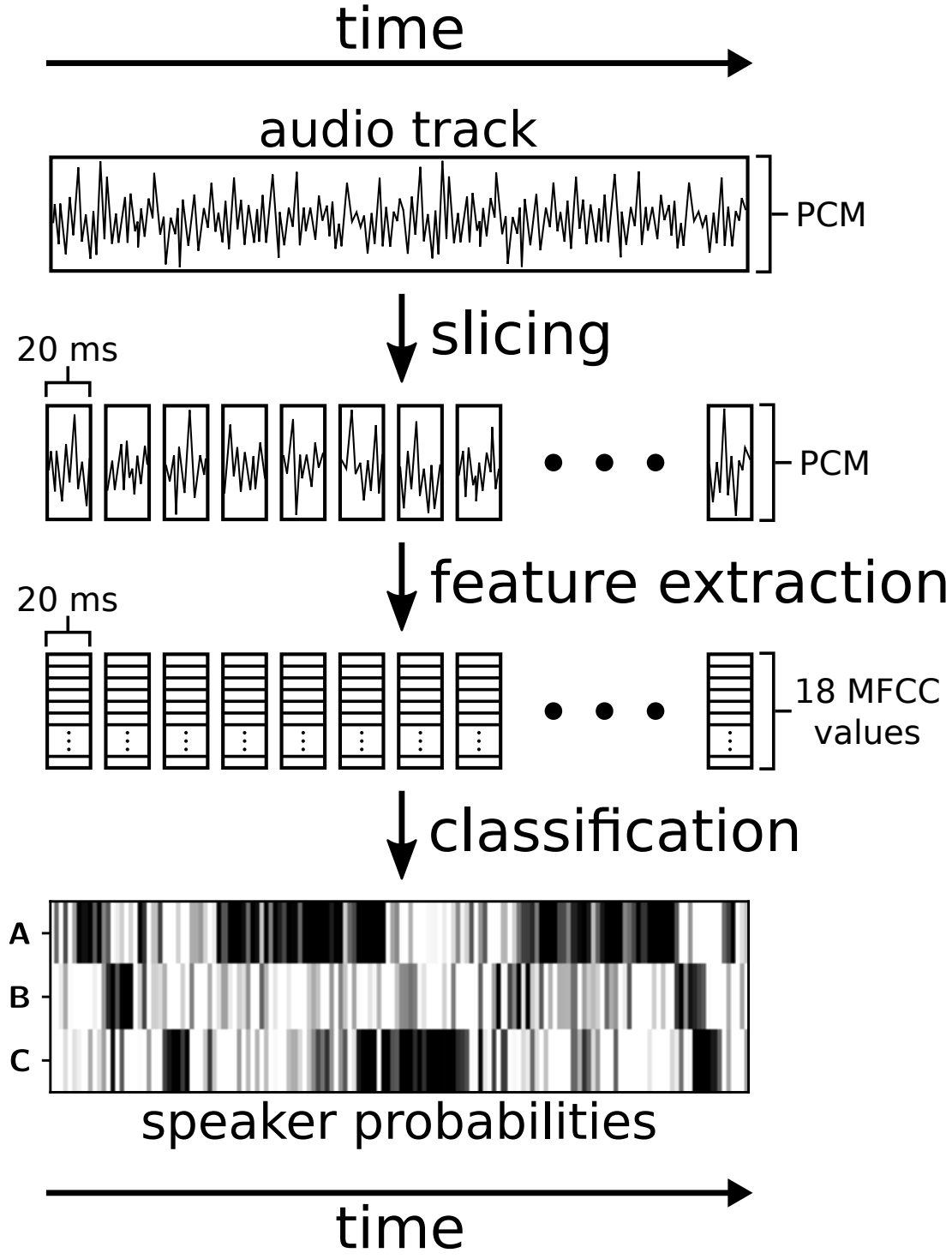
Puhujien tunnistaminen

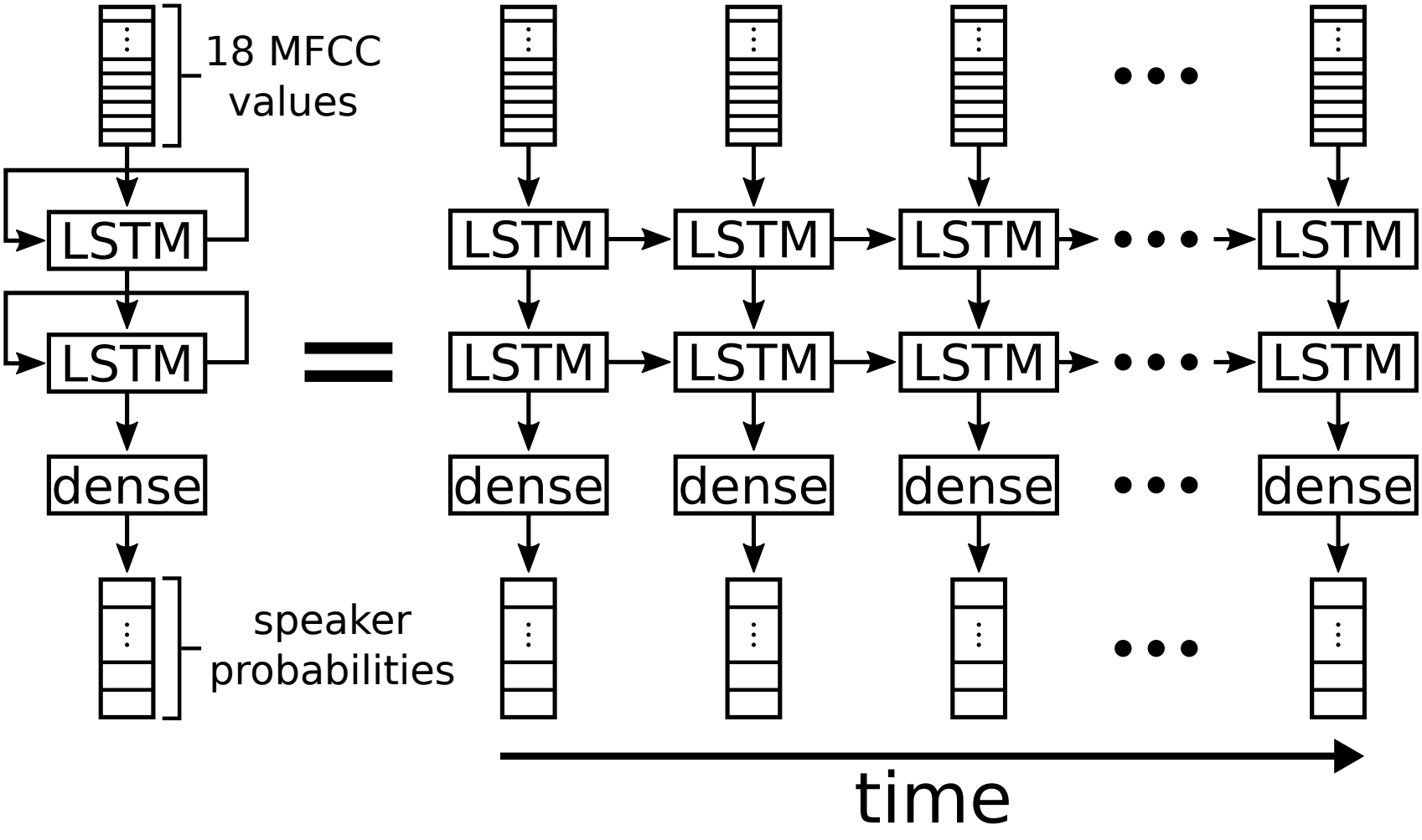
- Syötteenä on äänitiedosto keskustelusta.
- Ohjelma tunnistaa milloin kukin puhuja on äänessä.
- Ei välitetä puheen sisällöstä.

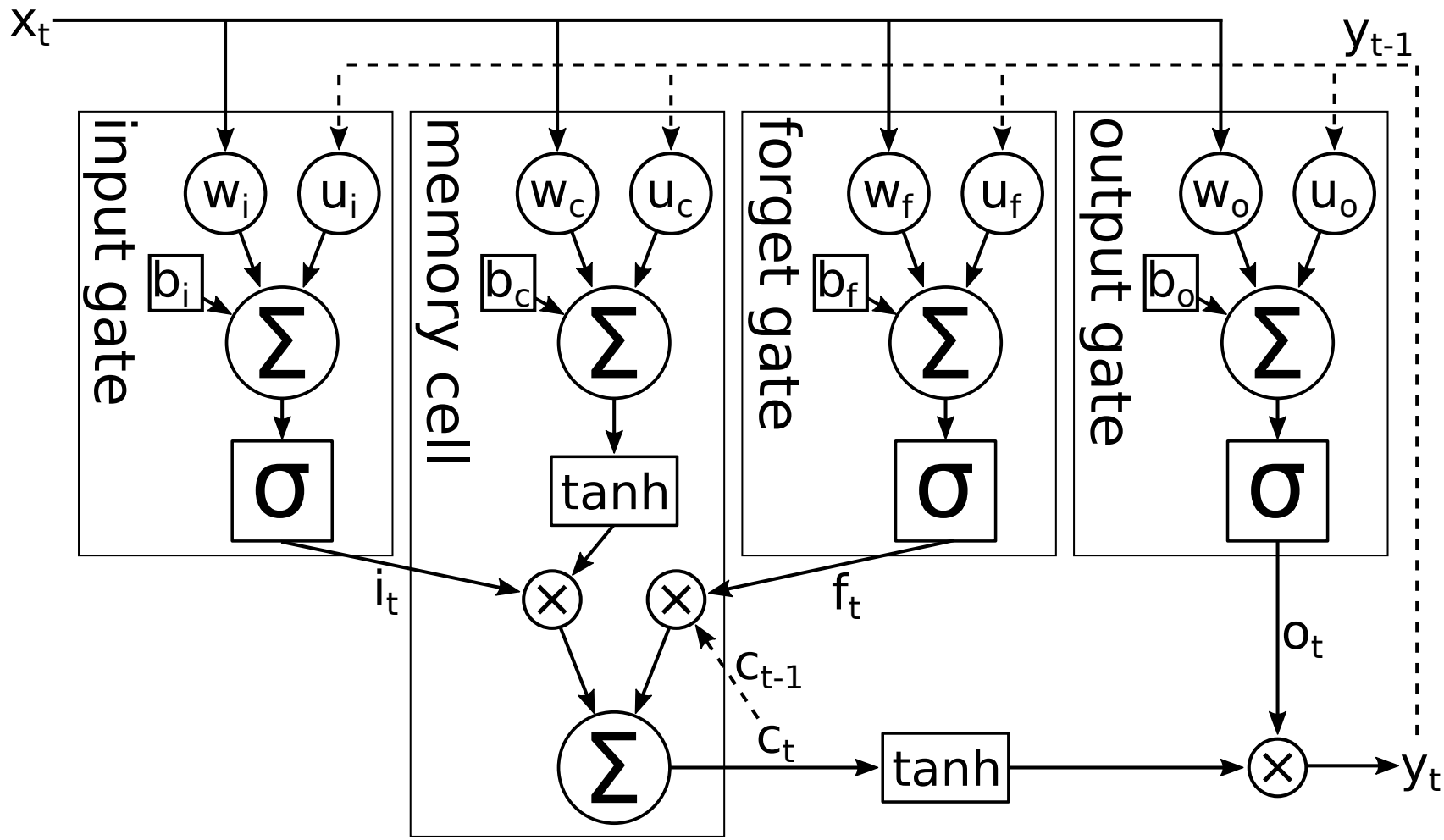


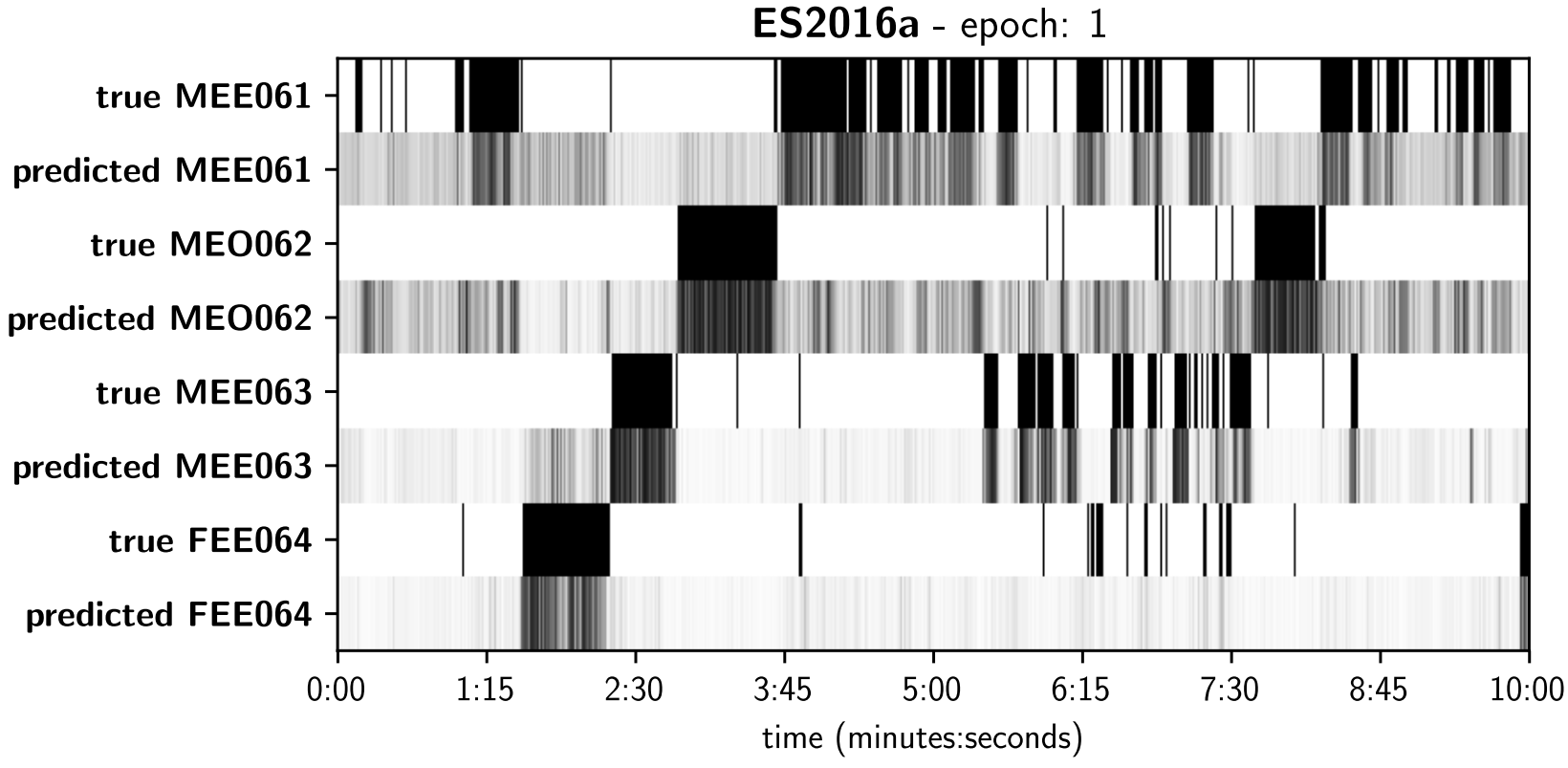
Toteutus

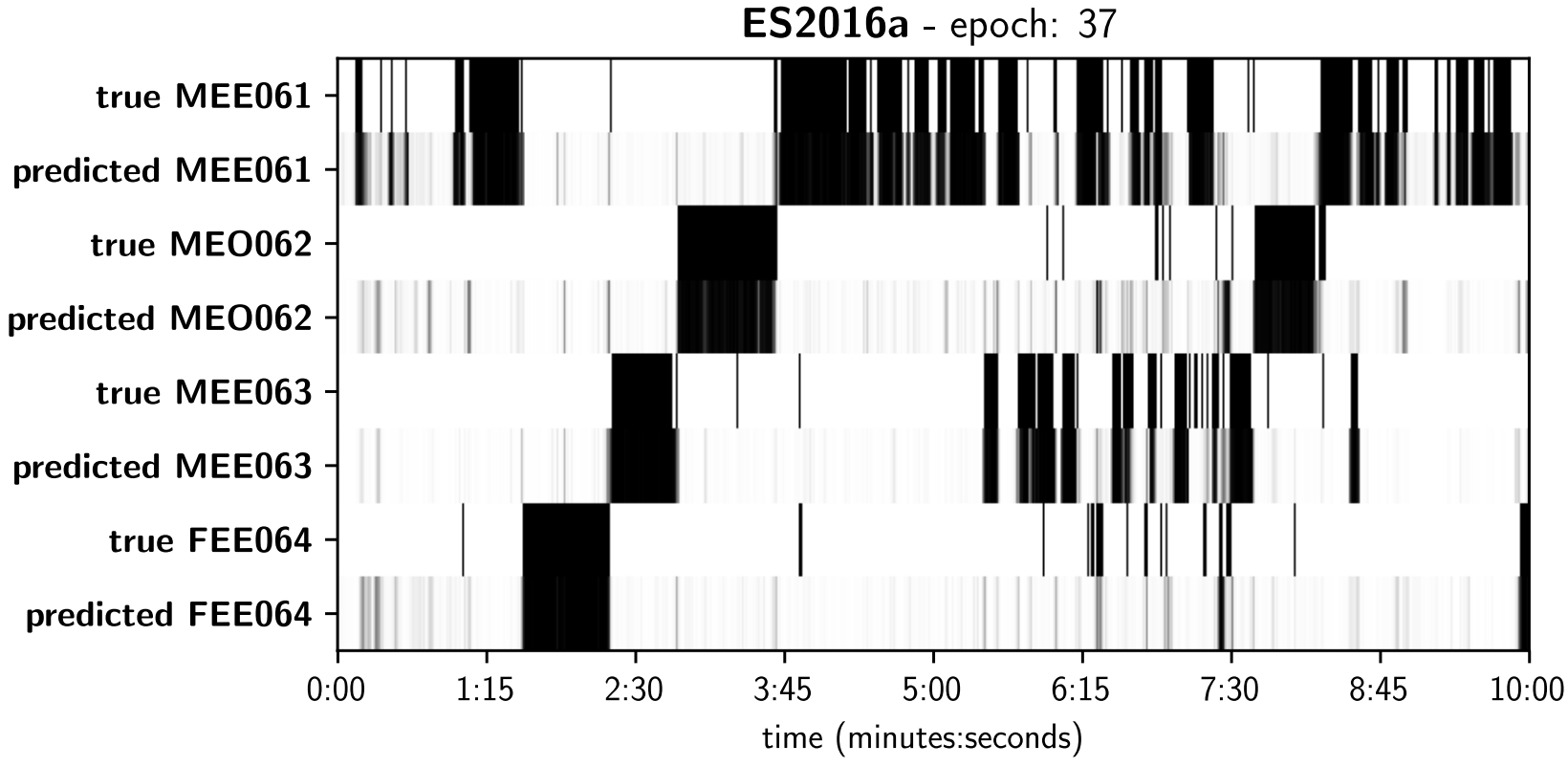
- Toteutus on tehty koneoppimistekniikoilla.
  - Luokitin opetetaan datan perusteella tekemään tunnistuksia.
- Äänen esiprosessointi helpommin käsiteltävään muotoon:
  - Mel-frequency cepstral coefficients (MFCC) -piirteet
- Takaisinkytkeytyvä neuroverkko käy ääntä läpi ja muistaa aiempia ajanhetkiä:
  - Long short-term memory (LSTM) -neuroverkkoluokitin


time
audio track
PCM
slicing
20 ms
PCM
feature extraction
20 ms
18 MFCC values
classification
A
B
C
speaker probabilities
time


18 MFCC values
LSTM
LSTM
dense
speaker probabilities
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
LSTM
dense
dense
dense
dense
time


x_t
y_{t-1}
input gate
memory cell
forget gate
output gate
w_i
u_i
b_i
w_c
u_c
b_c
w_f
u_f
b_f
w_o
u_o
b_o
Σ
σ
tanh
i_t
f_t
o_t
c_{t-1}
c_t
y_t


ES2016a - epoch: 1
true MEE061
predicted MEE061
true MEO062
predicted MEO062
true MEE063
predicted MEE063
true FEE064
predicted FEE064
0:00
1:15
2:30
3:45
5:00
6:15
7:30
8:45
10:00
time (minutes:seconds)

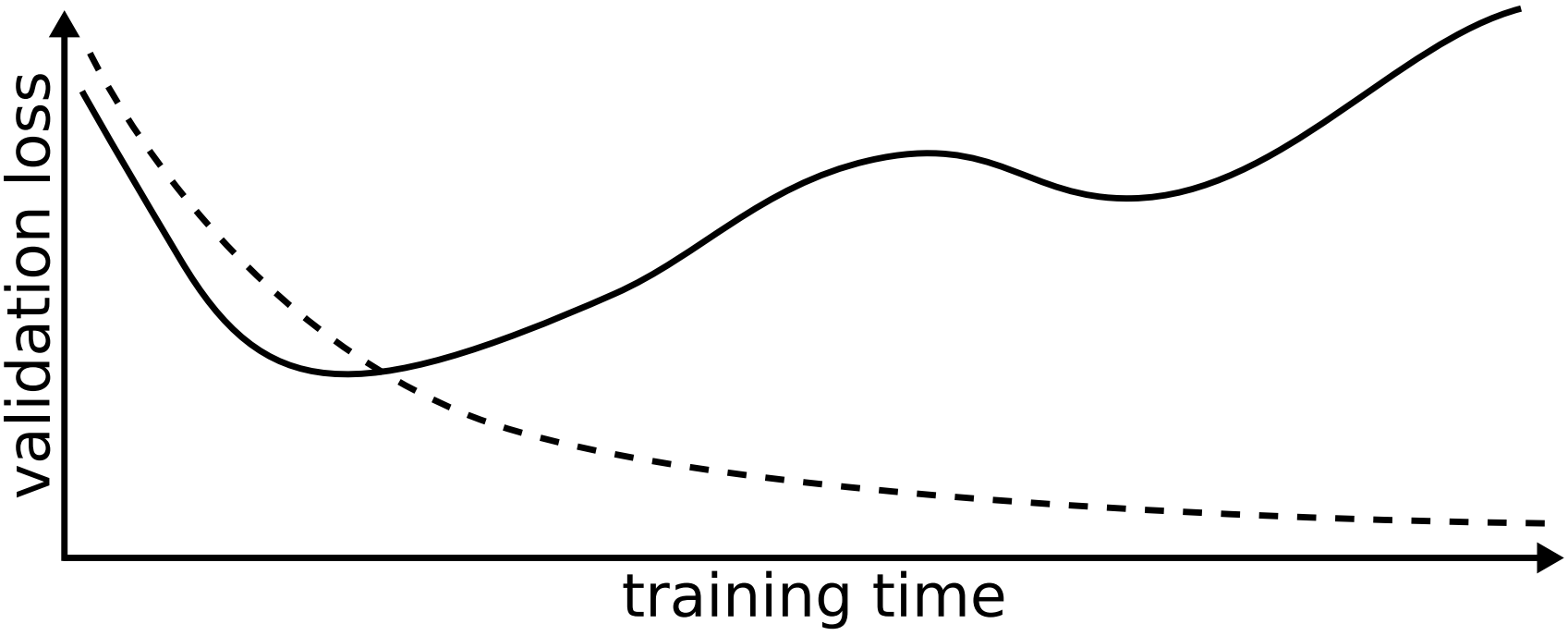

ES2016a - epoch: 37
true MEE061
predicted MEE061
true MEO062
predicted MEO062
true MEE063
predicted MEE063
true FEE064
predicted FEE064
0:00
1:15
2:30
3:45
5:00
6:15
7:30
8:45
10:00
time (minutes:seconds)

Hyperparametrit

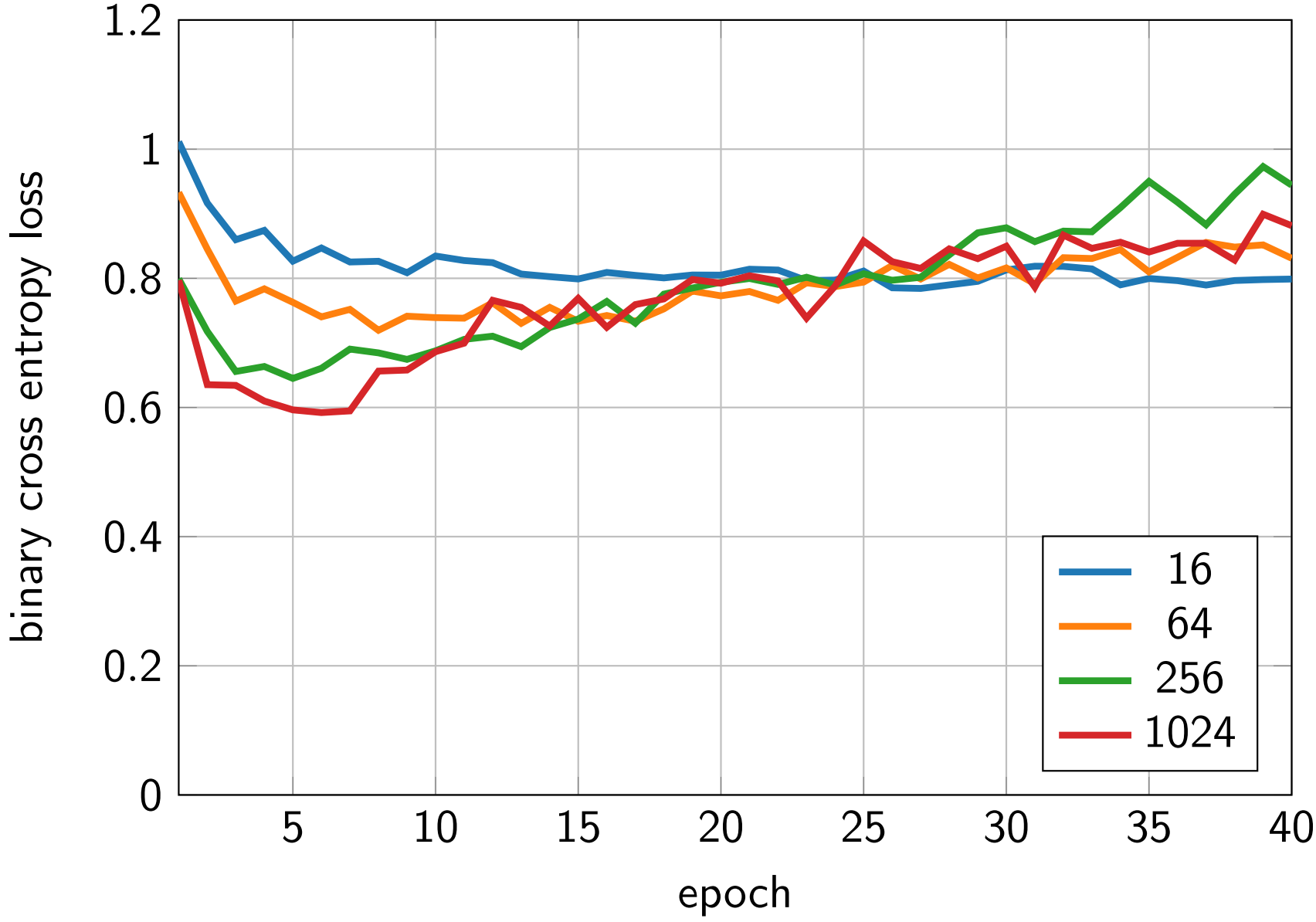
- Luokitinneuroverkolla on parametrejä, joita verkko ei opi automaattisesti.
- Säätämällä arvot oikein voi tuloksia saada paremmaksi.
- Tärkeimmät parametrit säätävät verkon oppimiskykyä.
  - Liian yksinkertainen verkko ei pysty oppimaan monimutkaisia tehtäviä.
  - Toisaalta liian suuri verkko ylioppii opetusdatalle, eikä yleisty uudelle datalle.
- Parametrejä:
  - LSTM-kerroksien lukumäärä (1, 2)
  - LSTM-kerrosten koko (16, 64, 256, 1024)
  - Dropout (0 %, 50 %)

Oppimiskäyrät ja ylioppiminen

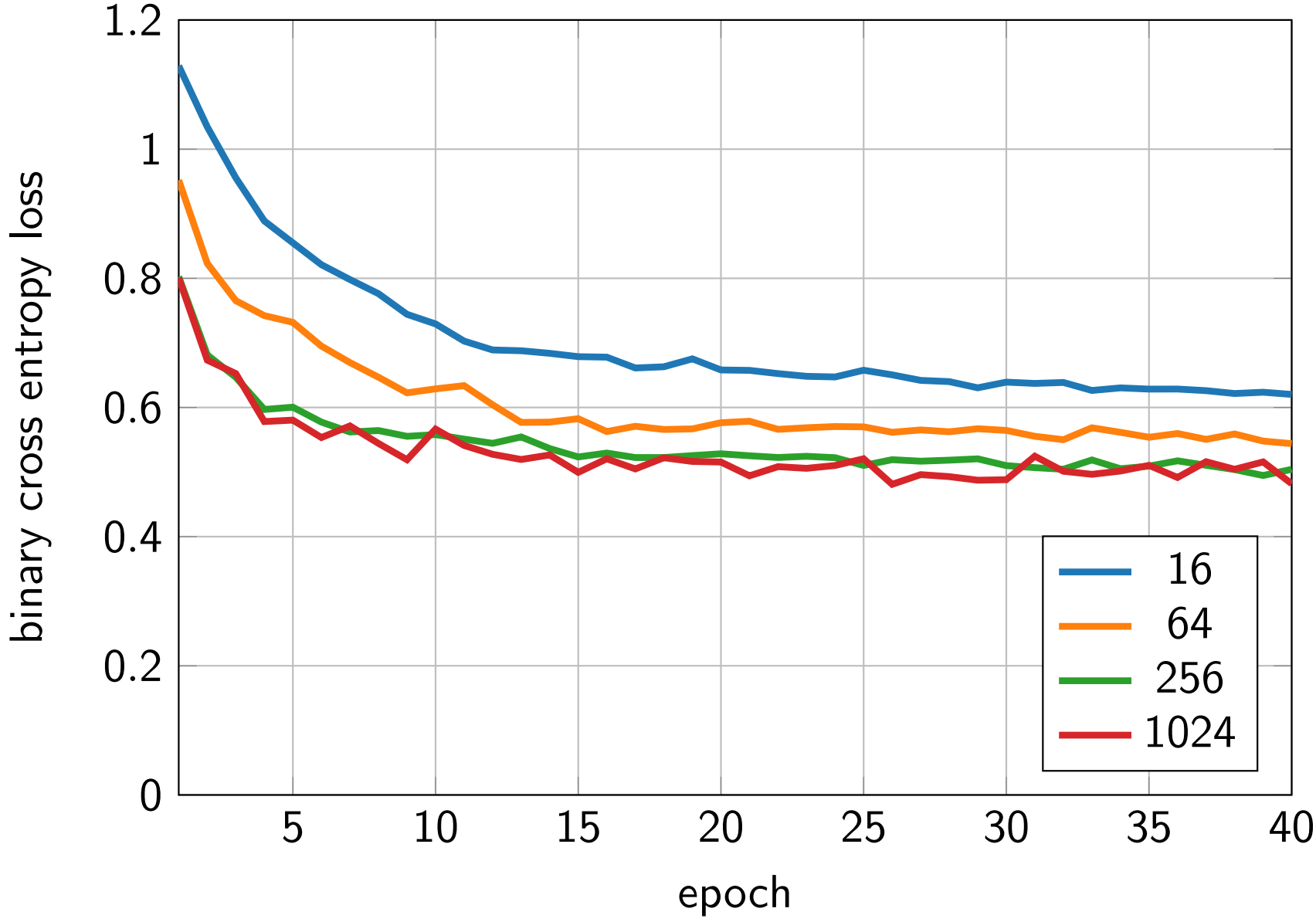
- Tavoitteena on saada luokittimen virhe mahdollisimman pieneksi.
- Virhe pienenee kun opetukseen käyttää enemmän aikaa.
- Jos luokitin alkaa ylioppia, niin virhe alkaakin kasvaa.



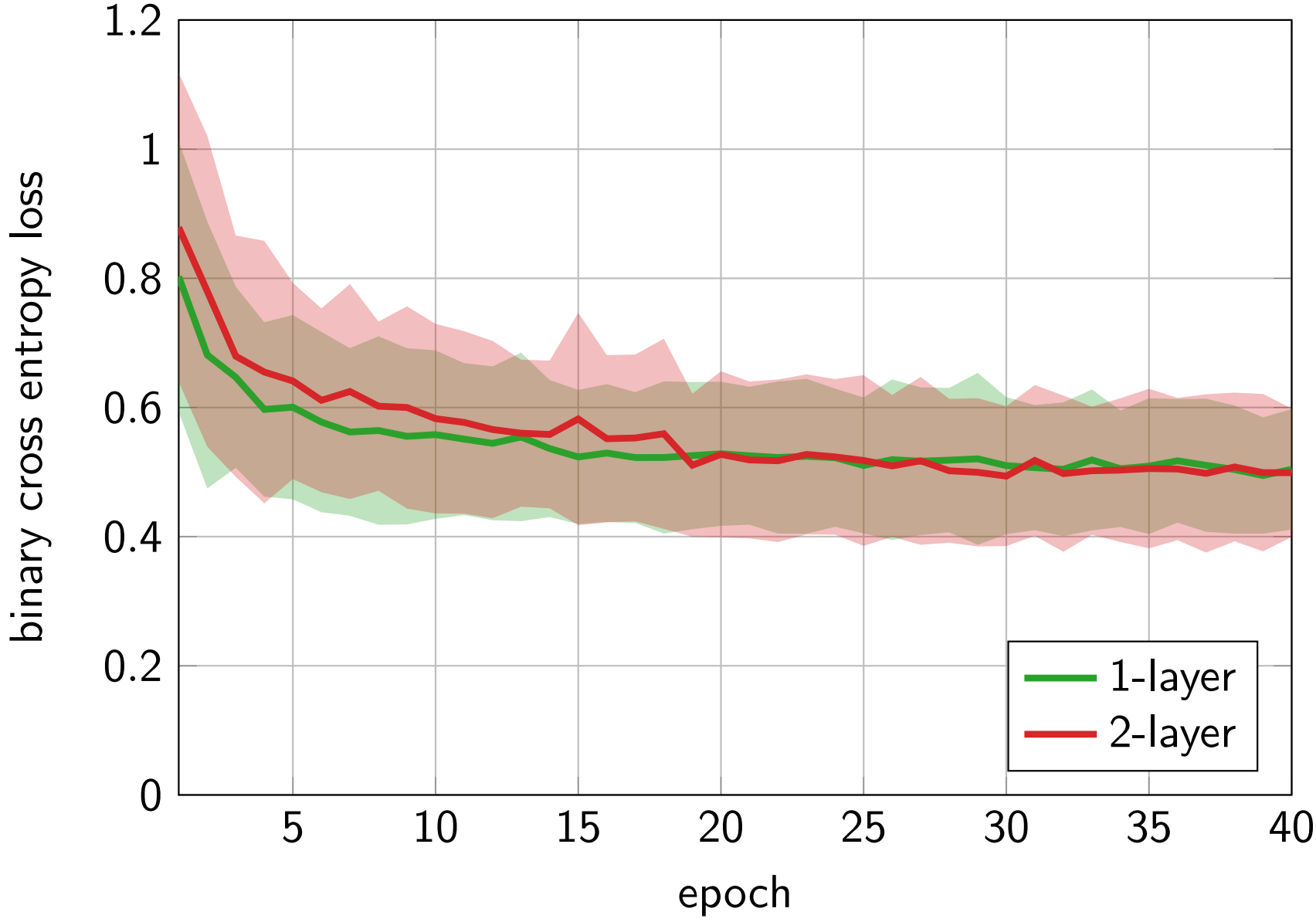
Verkon koon vaikutus



Dropout


1.2
1
0.8
0.6
0.4
0.2
0
binary cross entropy loss
5
10
15
20
25
30
35
40
epoch
16
64
256
1024

Kaksi LSTM-kerrosta

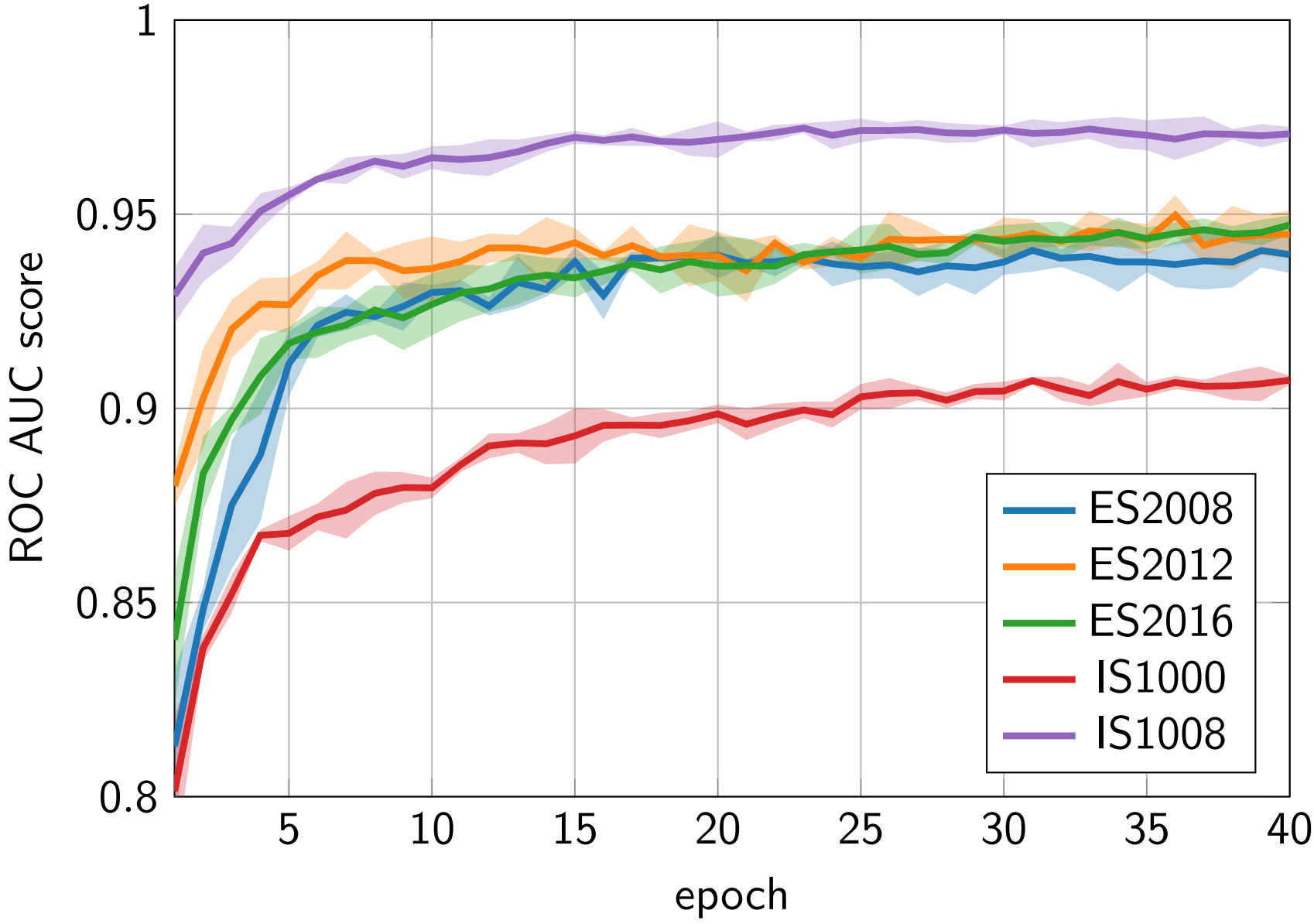


Opetukseen kuluva aika

training time per epoch (s)		
	1 layer	**2 layers**
layer size 16	35.1 ± 0.78	73.8 ± 3.50
layer size 64	36.7 ± 1.94	72.1 ± 2.68
layer size 256	35.0 ± 1.01	74.0 ± 1.84
layer size 1024	121.9 ± 5.22	340.1 ± 13.89

- Pienemmät verkot on nopeampi opettaa.

Tuloksia


1
0.95
0.9
0.85
0.8
ROC AUC score
5
10
15
20
25
30
35
40
epoch
ES2008
ES2012
ES2016
IS1000
IS1008