MATH 462

- problem defn
  Face Verification
- Important Fns.
  similarity S, $\mathbb{R}^4$
- Biz project

Lecture 16

3.11.2021

Face Verfication Problem (FV)

Given two images of { same object (person's face
                    { different object (two diff. person face)

Decide (Binary Class)

if they represent same +1
                  diff -1.

[Related classfication K people
  Given image x, which person is it?]

K-classification

Two types of generalization

(G1) m_1 photos of k different faces (balanced)

Training Set. Train class-model.

acc. p_{train}

m_2 photos of same k faces.

New photos what is the accuracy?

p_{test}

Gap $p_{train} - p_{test}$ generalization gap

Ex $.97 - .85 = .12$

model overfits if gap > 0.

Better model overfits less

$.91 - .89 = .02$

problem

F.V.

x image of a face

$p = (x^1, x^2)$ pair of images

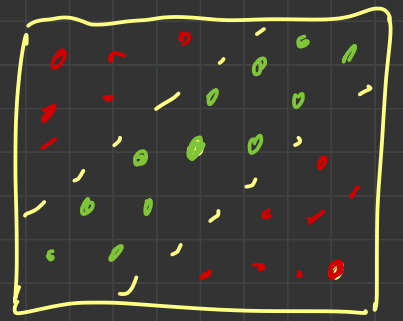
$y \in Y_\pm = \{-1, +1\}$

$$y(p) = \begin{cases} +1 & \text{similar pair (same face)} \\ -1 & \text{otherwise} \end{cases}$$

$S_m = \{(p_1, y_1), \dots, (p_m, y_m)\}$

$= \{(x_1^1, x_1^2, y_1), \dots, (x_m^1, x_m^2, y_m)\}$

GOAL Given $p = (x^1, x^2)$ classify as similar or not



$\mathcal{X}$

- diff picture of same face
- picture of diff. faces

Gl S_{train} most of pairs

S_{test}.

train model using S_{train} accuracy P_{train}

test model using S_{test} " " P_{test}

if $P_{train} - P_{test}$ too large. model overfits.

"Regularization" Deep Learning Book"
see chapter

→ Data Augmentation

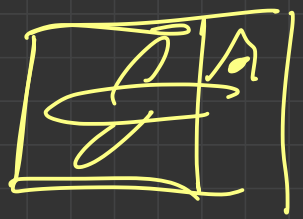
x replace with $\tilde{x} = T_i(x)$ for some transform.

T_i = add noise, crop, smooth,

change x, same image.

EG "learn features"

dog



G2 trained on Pceleb. out of dist gen.

test on reserved images from Pceleb.

Model

FV on other faces not in Pceleb

→ New objects. to verify never been seen before.

→ still faces.

→ not e.g. recogn. dog breeds.

what do the objects have in common?

In distribution generalization

G1 $p_{data}(x)$ S_m samples i.i.d.
independent, identically distributed.

generalization
usually p_{train} v.s. p_{test}

when S_{train} & S_{test}
each consisting of samples drawn i.i.d.
from same p_{data}.

G_2 out of distribution generation.

partially defined

$h(x)$ trained on $p_{data}(x)$

S_m i.i.d x from $p_{data}(x)$

$= \{(x_i, y_i)\}$ $y_i = y(x_i)$ true label

same starting point.

E.G. Classification.

Keep unknown the label for $y(x)$.

Change $p_{data}(x)$ to $p_{new}(x)$.

this part is not defined.?!!

G2 for FV:

new faces! people faces not in Ptrain

Works! Great!

So ood generalization is possible.

Research problem: define it in such a way cover cases works, not too broad.

→ Hard OOD gen.

→ easy OOD gen. (some confused)

sequel

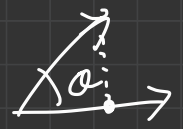
Background.

$$S^d = \{ v \in \mathbb{R}^d \mid \|v\| = 1 \}$$

Similarity

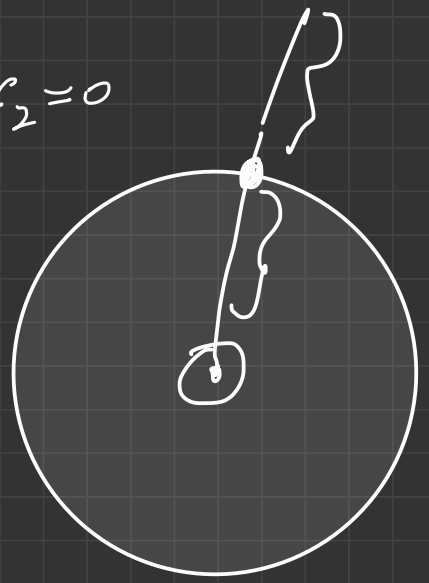
① Given $v_1, v_2 \in S^d$ $\quad v_1 \cdot v_2 = \cos\theta$

$\text{sim}(v_1, v_2) = \cos\theta$



② Given $f_1, f_2 \in \mathbb{R}^d$

$\cos\theta$ undefined if $f_1 \cdot f_2 = 0$



Defn

$$\text{sim}_\varepsilon : \mathbb{R}^d \times \mathbb{R}^d \to [-1, 1]$$

$$N_\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$$

$$N_\varepsilon(f) = \begin{cases} \frac{f}{\|f\|} & \text{if } \|f\| \geq \varepsilon \\ 0 & \text{o.w.} \end{cases}$$

$$\text{sim}_\varepsilon(f_1, f_2) = N_\varepsilon(f_1) \cdot N_\varepsilon(f_2)$$

$$sim_\varepsilon(f_1, f_2) = \begin{cases} \dfrac{f_1 \cdot f_2}{\|f_1\| \|f_2\|} & \text{if } \|f_1\| \geq \varepsilon \text{ and } \|f_2\| \geq \varepsilon \\ 0 & \text{if either of } \|f_1\| < \varepsilon \ \ \|f_2\| < \varepsilon \end{cases}$$

Method feature representation Learning

$$\mathcal{H} = \{ h(p) = s \mid h(p) = sim_\varepsilon(f(x_1', w), f(x^2, w)) - t \}$$

$f(x, w)$ DNN

$s \to s - t$ threshold score classifier.

$f: X \to \mathbb{R}^k$ features

$X \subseteq \mathbb{R}^d$

$p = (x^1, x^2)$

$x^1 \to f_1 = f(x^1, w)$

$x^2 \to f_2 = f(x^2, w)$

$s = sim_t(f_1, f_2)$

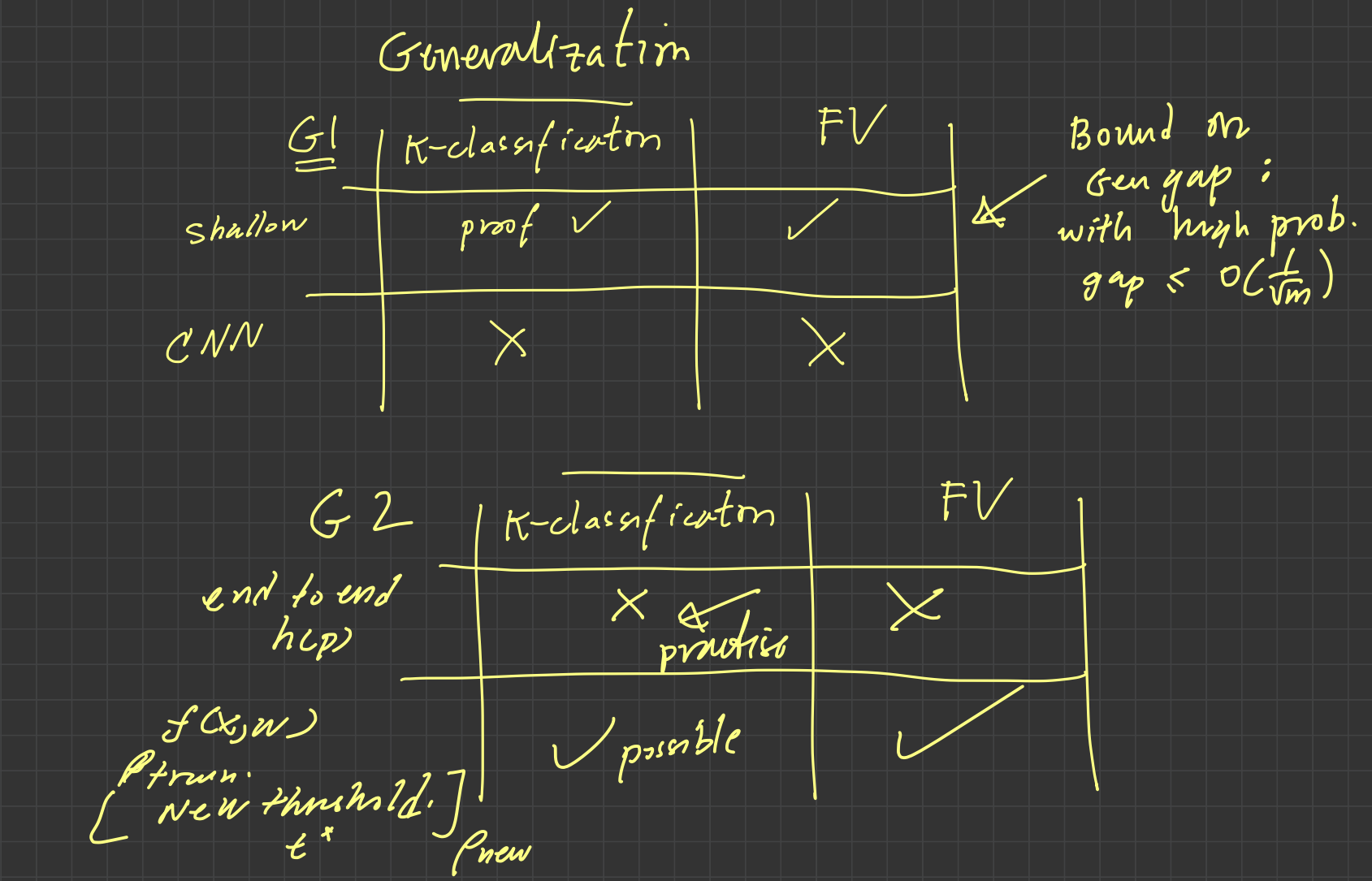
$h = s - t.$

$c(h) = sgn(h).$

Binary class same as before

$l_{class}(h, y) = l_{log}(h, y)$ or $l_{margin}(h, y)$

Loss $L(w, t) = \frac{1}{m} \sum_{i=1}^{m} l_{class}(h(p_i), y_i)$


Generalization
G1
K-classification
FV
shallow
proof ✓
✓
CNN
X
X
Bound on
Gen gap :
with high prob.
gap ≤ O(1/√m)
G2
K-classification
FV
end to end
hcp
X
X
practice
f(x,w)
Ptrain
[New threshold.
t*
✓ possible
✓
Pnew

train $f(x, w)$ on p_{train}.

using $\hat{L}(w, t)$

strip of the final layer (score classifier).

New dataset p_{new} different tasks.

① fine tune. where allow w to change
no proof. it can overfit.

② just change threshold.
i.e. only t will generalize.

$f(x)$.

model $h(w,x) = w \cdot f(x)$

$\hat{L}(w)$ (train loss.)

(train acc(h) $\leq \hat{L}(h(w,x))$)

Thm $L(w) = \mathop{\mathbb{E}}_{x \sim \rho} L(h(w,x), y)$ in theory

$\hat{L}_{test}(w)$ or $\frac{1}{m_1} \sum_{i=1}^{m} L(h(w,x_i), y_i)$ test loss.

Thm $|\hat{L}_{test} - \hat{L}_{train}| \leq \frac{C_\varepsilon}{\sqrt{m}} + \hat{\mathcal{R}}_{2m}(\mathcal{H})$

with prob $\geq 1-\varepsilon$

$f(x, w_0)$ ImageNet features.

Apply New data of vegetables.

New classifier

$h(x, \tilde{w}) = f(x, w_0) \cdot w$

freeze w_0, train w.

Acc Train, 93

Bound Test on veg.

$.93 - \frac{c_s}{\sqrt{m}}$ with high prob.

& enough train samples

Allow w_0 to train.

No Bound Acc Train .99