linear classifier

2024年9月12日 11:45

streches pixels → col. × respect spatial struc of img

1 General case:

ds $\{(x_i, y_i)\}_{i=1}^{N}$
(x_i: img, y_i: label index)

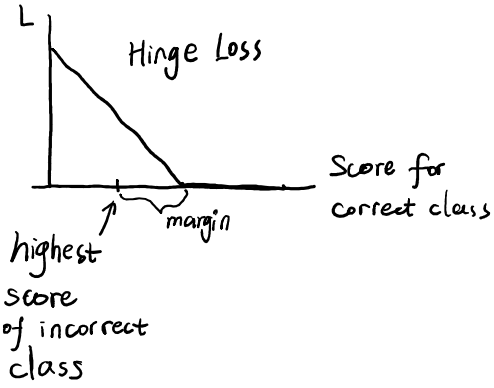
Loss L: how good our classifier is. $f(x, W) = Wx$

$$L = \frac{1}{N}\sum_i L_i(f(x_i, W), y_i)$$

($f(x_i, W)$: classifier score; y_i: lb index)

2 Multiclass SVM Loss

The score of the correct class should be higher than all the other scores.



Scores $s = f(x_i, W)$

Loss
$$L = \sum_{j \neq y_i} \begin{cases} 0 & \text{if } s_{y_i} \geq s_j + 1 \\ s_j - s_{y_i} + 1 & \text{o.w.} \end{cases}$$

$$= \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

($j \neq y_i$: not inc corr label)

e.g.

cat	**3.2**	1.3	2.2
car	5.1	**4.9**	2.5
frog	-1.7	2.0	**-3.1**

1° Compute the loss of a cat:

loop over all the incorrect classes

$L_{cat} = \max(0, 5.1 - 3.2 + 1)$

$L_{cat} = \max(0, 5.1 - 3.2 + 1)$
$+ \max(0, -1.7 - 3.2 + 1)$
$= 2.9 + 0$
$= 2.9$

$L_{car} = \max(0, 1.3 - 4.9 + 1)$
$+ \max(0, 2.0 - 4.9 + 1)$
$= 0$

$L_{frog} = \max(0, 2.2 - (-3.1) + 1)$
$+ \max(0, 2.5 - (-3.1) + 1)$
$= 6.3 + 6.6$
$= 12.9$

2° What happens to L_{car} if the scores for car img change a little bit?

Still 0. ∵ S_{car} a lot > any of other scores of incorrect classes

3° Min, max of loss?

min loss : 0 Correct cate has a s much higher than all incorrect cate.

max ∞ very ~ lower

4° If all the scores are random, what loss would we exp?

Supp: draw on scores from Gaussian dist. with very small σ,

⇒ s are small rand values

⇒ $E(s_j - s_{y_i}) \approx 0$

⇒ $\max(0, 0+1)$

1 per incorrect cate

⇒ $L_i = C - 1$

↑ cate ↑ correct

5° If sum over all the classes inc $i = y_i$?

All L to be inflated by 1.

多3 1 个 term $\max(0, s_{y_i} - s_{y_i} + 1)$

6° What if the loss used a mean instead of a Σ

Preference of the weight matrix remains the same.

7° What if we use $L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)^2$

Change scores in non-linear way

Cannot call it multi-class SVM loss, it shows dif pref on weight mat.

8° Sup some W with L=0. Unique?

No.

Original W:

= max(0, 1.3 - 4.9 + 1)
+max(0, 2.0 - 4.9 + 1)
= max(0, -2.6) + max(0, -1.9)
= 0 + 0
= 0

Using 2W instead:

= max(0, 2.6 - 9.8 + 1)
+max(0, 4.0 - 9.8 + 1)
= max(0, -6.2) + max(0, -4.8)
= 0 + 0
= 0

9° How should we choose between W and 2W if they both perform the same on the training data?

Other terms to eval pref on W.

Regularization: Beyond Training Error

$$L(W) = \frac{1}{N}\sum_{i=1}^{N} L_i(f(x_i, W), y_i) + \lambda R(W)$$

λ = regularization strength (hyperparameter)

Data loss: Model predictions should match training data

Regularization: Prevent the model from doing *too* well on training data

Simple examples

L2 regularization: $R(W) = \sum_k \sum_l W_{k,l}^2$ ← pref using more fea

L1 regularization: $R(W) = \sum_k \sum_l |W_{k,l}|$

Elastic net (L1 + L2): $R(W) = \sum_k \sum_l \beta W_{k,l}^2 + |W_{k,l}|$ ← pref weight on single fea

More complex:

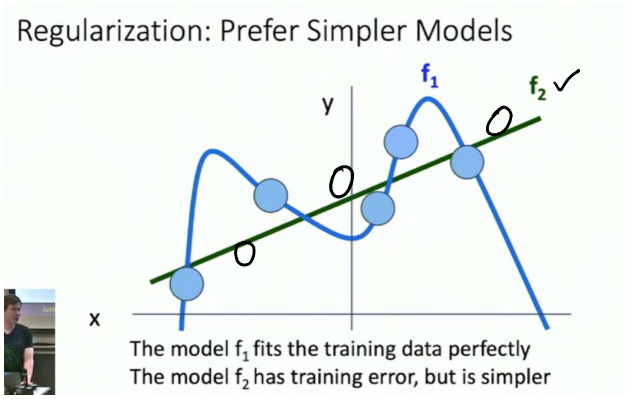
Dropout

Batch normalization

Cutout, Mixup, Stochastic depth, etc...

Purpose of Regularization:

- Express preferences in among models beyond "minimize training error"
- Avoid **overfitting**: Prefer simple models that generalize better
- Improve optimization by adding curvature

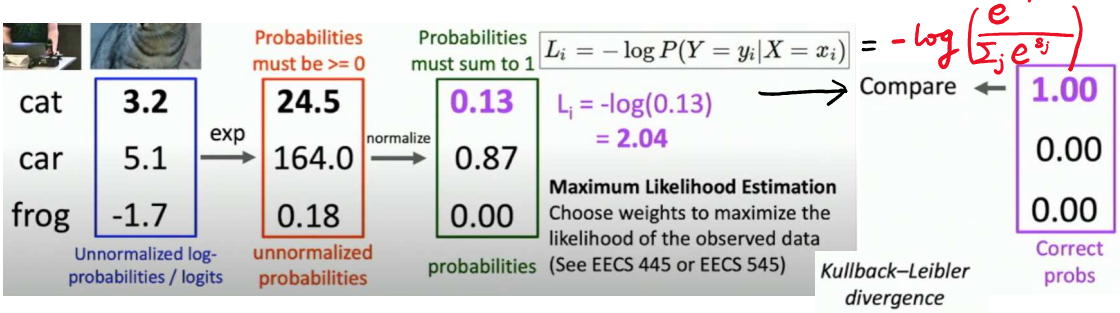


Regularization, 一 should usually use it.

3 Cross-Entropy Loss (Multinomial Logistic Regression)

raw scores → prob

$s = f(x_i; W)$ $\quad P(Y=k|X=x_i) = \frac{e^{s_k}}{\sum_j e^{s_j}}$



Kullback–Leibler divergence

$$D_{KL}(P\|Q) = \sum_y P(y)\log\frac{P(y)}{Q(y)}$$

1° possible L_i: min 0, max $+\infty$

2° all s: small rand values, L = ?

⇒ uniform in softmax

⇒ uniform prob $L = -\log(\frac{1}{C})$

① 训练开始时的 L 应该接近 $-\log(\frac{1}{C})$, 否则有大 bug

② 后续看到 越来越↑ 于 $-\log(\frac{1}{C})$, 有大 bug. 你的 classifier 比 random 还差

4 scores [10, -2, 3]

[10, 9, 9]

[10, -100, -100],

& $y_i = 0$

1° cross entropy loss ? >0

SVM $=0$

2° slightly change s of the last data pt ?

CE 会把 correct class 推向 $+\infty$, 反之 $-\infty$

SVM stay the same

3°

Q: What happens to each loss if I double the score of the correct class from 10 to 20?

A: Cross-entropy loss will decrease, SVM loss still 0