Lecture 3: Loss Functions and Optimization

Recall from last time: Challenges of recognition

Viewpoint Viewpoint

Illumination



This image is CC0 1.0 public domain

Deformation



This image by Umberto Salvagnin is licensed under CC-BY 2.0

Occlusion



This image by jonsson is licensed under CC-BY 2.0

Clutter



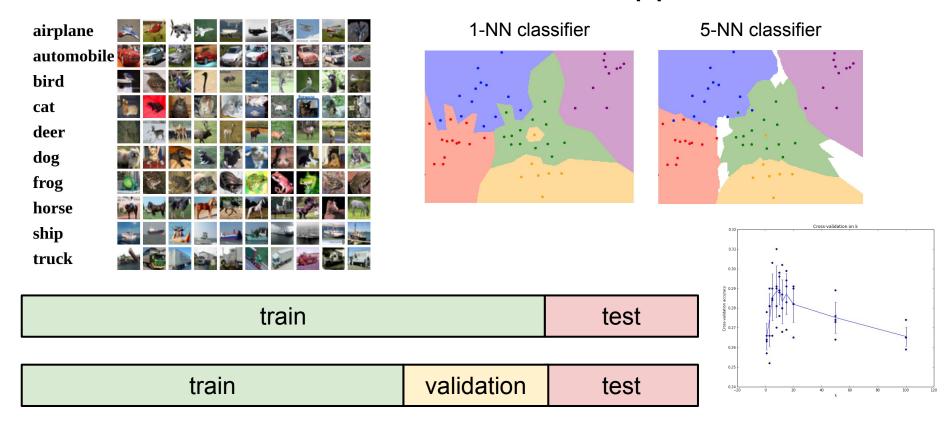
This image is CC0 1.0 public domain

Intraclass Variation

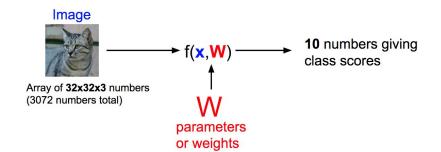


This image is CC0 1.0 public domain

Recall from last time: data-driven approach, kNN



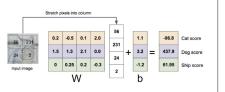
Recall from last time: Linear Classifier



$$f(x,W) = Wx + b$$

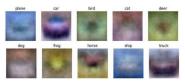


$$f(x,W) = Wx$$



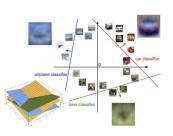
Visual Viewpoint

One template per class



Geometric Viewpoint

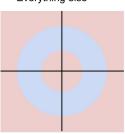
Hyperplanes cutting up space



Class 1:

Class 2: Everything else

1 <= L2 norm <= 2

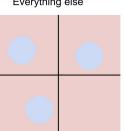


Class 1:

Three modes

Class 2:

Everything else



Recall from last time: Linear Classifier







airplane	-3.45	-0.51	3.42
automobile	-8.87	6.04	4.64
bird	0.09	5.31	2.65
cat	2.9	-4.22	5.1
deer	4.48	-4.19	2.64
dog	8.02	3.58	5.55
frog	3.78	4.49	-4.34
horse	1.06	-4.37	-1.5
ship	-0.36	-2.09	-4.79
truck	-0.72	-2.93	6.14

TODO:

- Define a loss function that quantifies our unhappiness with the scores across the training data.
- Come up with a way of efficiently finding the parameters that minimize the loss function. (optimization)

Cat image by Nikita is licensed under CC-BY 2.0; Car image is CC0 1.0 public domain; Frog image is in the public domain

Suppose: 3 training examples, 3 classes.

With some W the scores f(x, W) = Wx are:

	1	4
Á		3
		7
		7-





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

frog

Suppose: 3 training examples, 3 classes.

With some W the scores f(x, W) = Wx are:

 $\{(x_i, y_i)\}_{i=1}^N$

Where x_i is image and y_i is (integer) label

A **loss function** tells how

good our current classifier is

Given a dataset of examples

Loss over the dataset is a

sum of loss over examples: $L = \frac{1}{N} \sum_{i} L_i(f(x_i, W), y_i)$

5.1

-1.7





3.2 cat

car

frog

2.2

-3.1

Fei-Fei Li & Justin Johnson & Serena Yeung







cat

3.2

1.3

2.2

5.1 car

4.9

-1.7 frog

2.0

2.5

-3.1

Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s = f(x_i, W)$

the SVM loss has the form:

$$L_i = \sum_{j \neq y_i} \begin{cases} 0 & \text{if } s_{y_i} \geq s_j + 1 \\ s_j - s_{y_i} + 1 & \text{otherwise} \end{cases}$$
$$= \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

Suppose: 3 training examples, 3 classes.

With some W the scores f(x, W) = Wx are:







cat

car

frog

3.2

5.1

-1.7

1.3

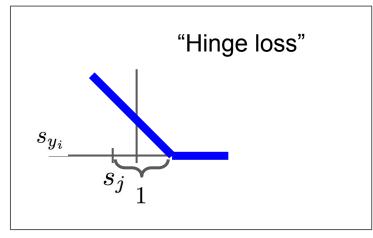
2.2

4.9

2.0

-3.1

Multiclass SVM loss:



2.5
$$L_i = \sum_{j \neq y_i} \begin{cases} 0 & \text{if } s_{y_i} \geq s_j + 1 \\ s_j - s_{y_i} + 1 & \text{otherwise} \end{cases}$$

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







cat

3.2

1.3

2.2

car 5.1

4.9

2.5

frog -1.7

.7 2.0

-3.1

Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s = f(x_i, W)$

the SVM loss has the form:

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



3.2





Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s=f(x_i,W)$

cat

car

5.1

frog -1.7

Losses: 2.9

1.3

1.32.24.92.5

2.0

-3.1

the SVM loss has the form:

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

 $= \max(0, 5.1 - 3.2 + 1)$

 $+\max(0, -1.7 - 3.2 + 1)$

 $= \max(0, 2.9) + \max(0, -3.9)$

= 2.9 + 0

= 2.9







Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

Multiclass SVM loss:

and using the shorthand for the scores vector: $s = f(x_i, W)$

3.2 1.3 cat

5.1

4.9

2.2

2.0

2.5

-1.7 frog Losses:

car

2.9

-3.1

the SVM loss has the form:

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

 $= \max(0, 1.3 - 4.9 + 1)$

 $+\max(0, 2.0 - 4.9 + 1)$

 $= \max(0, -2.6) + \max(0, -1.9)$

= 0 + 0

= 0



cat

Losses:





12.9

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

Multiclass SVM loss:

and using the shorthand for the scores vector: $s = f(x_i, W)$

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

the SVM loss has the form:

 $L_i = \sum_{j
eq y_i} \max(0, s_j - s_{y_i} + 1)$ $= \max(0, 2.2 - (-3.1) + 1)$

 $+\max(0, 2.5 - (-3.1) + 1)$ $= \max(0, 6.3) + \max(0, 6.6)$

= 6.3 + 6.6

= 12.9

2.9







2.5

-3.1

Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s = f(x_i, W)$

the SVM loss has the form: 2.2 1.3

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

Loss over full dataset is average:

 $L = \frac{1}{N} \sum_{i=1}^{N} L_i$

L = (2.9 + 0 + 12.9)/3= 5.27

12.9

4.9

2.0

5.1

-1.7

2.9

cat

car

frog

Losses:

Fei-Fei Li & Justin Johnson & Serena Yeung

Lecture 3 - 20 April 10, 2018



5.1

-1.7

2.9





Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s=f(x_i,W)$

the SVM loss has the form:

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

Q: What happens to loss if car scores change a bit?

cat **3.2**

car

frog

Losses:

1.3

2.2

12.9

4.92.52.0-3.1

Fei-Fei Li & Justin Johnson & Serena Yeung

April 10, 2018









3.2 cat

1.3

2.2 2.5

5.1 car

4.9 2.0

-3.1

-1.7 frog 2.9 Losses:

12.9

Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s = f(x_i, W)$

the SVM loss has the form:

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

Q2: what is the min/max possible loss?







Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s = f(x_i, W)$

the SVM loss has the form:

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

Q3: At initialization W is small so all $s \approx 0$. What is the loss?

3.2 cat

car

frog

Losses:

1.3

2.2 2.5

4.9 5.1 -1.7 2.0

-3.1

2.9

12.9







Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s=f(x_i,W)$

the SVM loss has the form:

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

Q4: What if the sum was over all classes? (including j = y_i)

cat **3.2**

car

1.3

2.2 2.5

5.1 **4.9**

2.0 **-3.1**

frog -1.7 Losses: 2.9

12.9



-1.7





3.2 cat

1.3

2.2

5.1 car

4.9 2.0

2.5 -3.1

frog 2.9 Losses:

12.9

Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s = f(x_i, W)$

the SVM loss has the form:

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

Q5: What if we used mean instead of sum?





1.3

4.9



2.2

2.5

12.9

frog -1.7 2.0 **-3.1**

2.9

3.2

5.1

cat

car

Losses:

.9 0

..

Multiclass SVM loss:

Given an example (x_i, y_i) where x_i is the image and where y_i is the (integer) label,

and using the shorthand for the scores vector: $s=f(x_i,W)$

the SVM loss has the form:

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

Q6: What if we used

$$L_i = \sum_{j
eq y_i} \max(0, s_j - s_{y_i} + 1)^2$$