Midterm I Review

Katie DC

Topics to Review

- Topics:
  - Regression, clustering, classification, convolution, linear filtering, derivative filters, Bayes filter and its implementations, histogram filter, particle filter, Monte Carlo localization, grid localization, classical recognition, bag of words, shallow and deep neural networks
- Background material
  - Basic linear algebra, calculus, and convolution properties
- Algorithms and methods covered in lecture
  - PageRank, SVM, kNN, k-means, Backpropagation, filtering implementations, deep neural network implementations
- Topics covered in the MPs
  - Image manipulation, CV implementations
- More topics: Questions will be discussion, analysis, and problem solving

Convolutions

$f = [\,3 \;\; 1 \;\; 2\,]$

$g = [\,3 \;\; 2 \;\; 1\,]$

What is the convolution of f and g?

$$\underline{f * g[n]} = \sum_{m=-\infty}^{\infty} f[m]\, g[n-m]$$

$f*g[0] = 3 \cdot 3 = 9$

$f*g[1] = 3 \cdot 2 + 3 \cdot 1 = 9$

$f*g[2] = 3 \cdot 1 + 1 \cdot 2 + 2 \cdot 3 = 11$

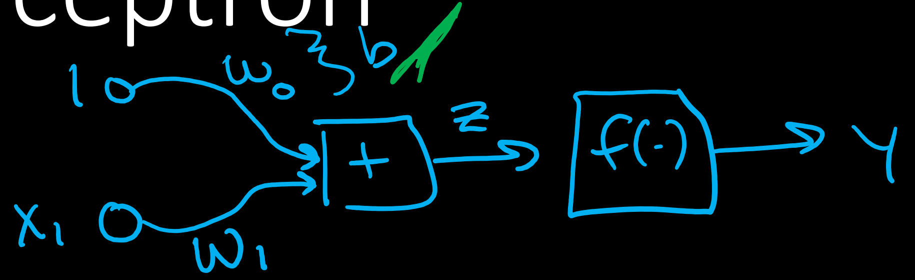
$f*g[3] = 1 \cdot 1 + 2 \cdot 2 = 5$

$f*g[4] = 1 \cdot 2 = 2$

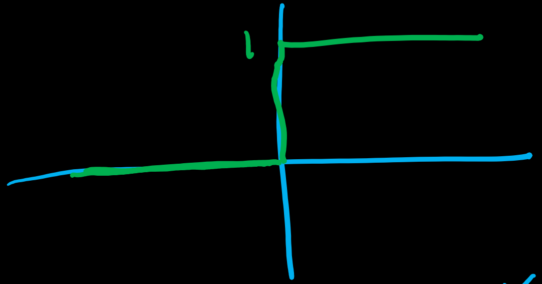
$f*g[5] = 0$

k	-2	-1	0	1	2	3
f			3	1	2	
g[-k]	1	2	3			
g[1-k]		1	2	3		
g[2-k]			1	2	3	

Perceptron



activation



$f(z) = \mathrm{sgn}(z)$

$x = [1 \quad x_1]^T$

$w = [w_0 \quad w_1]^T$

$z = w^T x$

$y = f(w^T x) = \mathrm{sgn}(w^T x)$

x	y_d
1	1
5	0

input	weights	$w^T x$	$\hat{y}$	
[1 1]	[.5 .5]	1	1	✓
[1 5]	"	3	1	✗

update rules: $w_{k+1} = w_k$ if okay!

$w_{k+1} = w_k - \eta x_k$ if $\hat{y} \geq 0$ when c_2

$w_{k+1} = w_k + \eta x_k$ if $\hat{y} < 0$ when c_1

perceptron convergence alg.

Gradient Descent Update

consider a perceptron w/ one output, one input, and a linear activation:

$$y = f(w^T x) = w^T x = [w_0 \quad w_1] \begin{bmatrix} 1 \\ x_1 \end{bmatrix}$$

$$= w_0 + w_1 x$$

Let's consider the optimization of the error:

$$E(w) = \tfrac{1}{2} e^2, \quad e = y_d - y, \quad \text{at sample } \begin{matrix} x = 1 \\ y_d = 2 \end{matrix}$$

$$= \tfrac{1}{2}(y_d - y)^2$$

$$= \tfrac{1}{2}(y_d - w_0 - w_1 x)^2$$

$$= \tfrac{1}{2}(2 - w_0 - w_1)^2$$

Local Gradient Descent

Now let's think about an MLP, w/ activation

$$f(x) = \frac{1}{1+e^{-x}}$$

$$f'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})(1+e^{-x})} = \overbrace{\frac{1}{1+e^{-x}}}^{f(x)} \; \overbrace{\frac{e^{-x}}{1+e^{-x}}}^{1-f(x)}$$

$$1 - \frac{1}{1+e^{-x}} = \frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-x}}$$

$$f'(x) = f(x)[1 - f(x)]$$

Backprop update equations:

for output unit j

$$\delta_j = e_j \underbrace{f_j'(\cdot)}_{f(\cdot)=y_j} = (y_{d_j} - y_j)\, y_j (1 - y_j)$$

for hidden unit i

$$\delta_i = f_i'(\cdot) \sum_j \delta_j w_{ji} = y_i (1 - y_i) \sum_j \delta_j w_{ji}$$

Bayesian Filter (1)

estimate the state of a door
↳ open or closed
o c

sensor model

$bel(x_0 = o) = .5$
$bel(x_0 = c) = .5$

$p(z_t = so \mid x_t = is\,o) = .6$
$p(z_t = sc \mid x_t = is\,o) = .4$
$p(z_t = so \mid x_t = is\,c) = .2$
$p(z_t = sc \mid x_t = is\,c) = .8$

suppose our agent can push door
$u_t = 1$ when pushing, $u_t = 0$ when do nothing

$p(x_t = is\,o \mid u_t = 1, x_{t-1} = is\,o) = 1$
$p(x_t = is\,c \mid u_t = 1, x_{t-1} = is\,o) = 0$
$p(x_t = is\,o \mid u_t = 1, x_{t-1} = is\,c) = .8$
$p(x_t = is\,c \mid u_t = 1, x_{t-1} = is\,c) = .2$

Bayesian Filter (2)

suppose we first do nothing

$$\overline{bel}(x_1) = \sum_{x_0} p(x_1 \mid u_1, x_0)\, bel(x_0)$$

$$= p(x_1 \mid u_1 = 0, x_0 = iso) \cdot bel(x_0 = iso)$$

$$+ p(x_1 \mid u_1 = 0, x_0 = isc)\, bel(x_0 = isc)$$

$$\overline{bel}(x_1 = iso) = 1 \cdot .5 + 0 \cdot .5 = .5$$

$$\overline{bel}(x_1 = isc) = .5$$

Bayesian Filter (3)

Recall Bayesian Filter Alg:

$bel(x_1) = \eta\, p(z_1 = so \mid x_1) \cdot \overline{bel}(x_1)$

$bel(x_1 = is\,o) = \eta \cdot .6 \cdot .5 = \eta \cdot .3 = .75$

$bel(x_1 = is\,c) = \eta \cdot .2 \cdot .5 = \eta \cdot .1 = .25$

$\eta = (.3 + .1)^{-1} = 2.5$

$u_2 = 1,\ z_2 = so$

$\overline{bel}(x_2 = is\,o) = 1 \cdot .75 + .8 \cdot .25 = .95$

$\overline{bel}(x_2 = is\,c) = 0 \cdot .75 + .2 \cdot .25 = .05$

$bel(x_2 = is\,o) = \eta \cdot .6 \cdot .95 \approx .983$

$bel(x_2 = is\,c) = \eta \cdot .2 \cdot .05 \approx .017$

Short Answer Questions?

- Suppose you want to design a ConvNet for image classification
  - What are the different layering components in ConvNets?
  - About how many parameters are there in such networks (order of magnitude)?
  - Can you name a dataset that you might use? What are some of the nice properties about benchmark datasets?
- Think about the classifier you trained in MP2. What are some of the concerns you have about using it on a system out in the real-world?
- For images, what is the difference between different types of image noise?
- Describe the SVM optimization. What are support vectors?