Machine Learning CS 4641

Data Analysis Toolbox 2

Nimisha Roy

Lecturer, SCI

Director of Online Undergraduate Initiatives

College of Computing

Slides adapted from Max Roozbahani, Rod Borela

Will cover

- Image Kernels and filtering
- Probability
- Information Theory
- Linear Algebra
- Optimization
- KMeans

Kernels

Convolution

Kernel to detect vertical edge

perform element wise operation

1	0	-1
1	0	-1
1	0	-1

1 example of kernel

$$= \quad 0 \times 0 + -1 \times 1 + 0 \times 2 + -1 \times 1 = -2$$

0	1	2	4	5	6	0	2
2	1	3	8	9	255	72	83
1	2	1	4	79	65	53	33
43	97	15	67	104	77	163	43
36	173	13	76	205	89	174	34
63	163	113	86	209	91	185	98
84	153	123	96	134	101	196	121
96	143	133	79	135	103	216	211

← image

-2							

filtered image

Kernel = [1 0 -1 1 0 -1 1 0 -1]

Subimage = [0 1 2 2 1 3 1 2 1]

→ dot product

Multivariate Gaussian Distribution $\quad d$ dimensions

$\mu \; (d \times 1)$

$$\Sigma \; (d \times d) = \begin{array}{c} \begin{array}{cc} \sigma_x^2 & \sigma_{xy} \end{array} \\ \begin{bmatrix} 1 & -20 \\ -20 & 625 \end{bmatrix} \\ \begin{array}{cc} \sigma_{yx} & \sigma_y^2 \end{array} \end{array}$$

$$\Sigma = \frac{\bar{X}^T \bar{X}}{N}$$

drawback — not scaled

$$Cor = \frac{\bar{X}^{*T} \bar{X}^{*}}{N}$$

$$\bar{X}^{*} = \frac{X - \mu}{\sigma}$$

scaled between 0 & 1

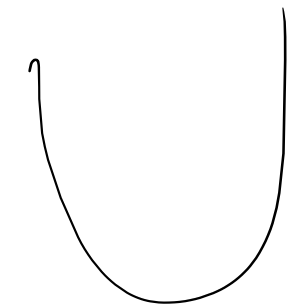
Optimization

$$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x - y \\ 2y - x \end{bmatrix}$$

$$f(x,y) = x^2 + y^2 - xy$$

Unconstrained Optimization problem

$$\text{Hessian Matrix} = \begin{bmatrix} \partial^2 f/\partial x^2 & \partial^2 f/\partial x \partial y \\ \partial^2 f/\partial y \partial x & \partial^2 f/\partial y^2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$



$$|\text{Hessian}| = 2 \times 2 - (-1 \times -1) = 3 > 0 \rightarrow \text{convex}$$

$$\frac{\partial f(x,y)}{\partial x} = 0 \Rightarrow 2x - y = 0 \Rightarrow x = y/2$$

$$(x^*, y^*) = (0, 0)$$

$$\frac{\partial f(x,y)}{\partial y} = 0 \Rightarrow 2y - x = 0 \Rightarrow y = x/2$$

$g(x,y) = 1-x-y$ $\quad g(x,y) \le 0$ Inequality constraint Optimization

$L(x,y,\lambda) = x^2+y^2-xy + \lambda(1-x-y)$ $\quad \lambda \ge 0$

$\frac{\partial L}{\partial x} = 0 \quad \frac{\partial L}{\partial y} = 0$

$2x-y-\lambda = 0$

$2y-x-\lambda = 0$

$\lambda = 2-3y$
$= 2-3\times\frac{1}{2}$
$= 0.5$

$\lambda \cdot g(x,y) = 0$

Case 1 $\quad \lambda = 0, \quad g(x,y) < 0$

$2x-y=0 \quad 2y-x=0$

$x^*, y^* = (0,0)$

$g(x,y) = 1-0-0$
$= 1 > 0$

Case 2: $\lambda > 0, \quad g(x,y) = 0$

$1-x-y = 0$ $\quad (x^*,y^*) = (0.5, 0.5)$

$1 = x+y$

$x = 1-y = 1-\frac{1}{2} = \frac{1}{2}$

$2(1-y)-y = \lambda$

$2-2y-y = \lambda$

$2-3y = \lambda$

$2-3y = 3y-1$
$6y = 3$
$y = 1/2$

$2y-1+y-\lambda = 0$

$3y-1 = \lambda$

Georgia Tech

KMeans

Mean of distances
L_2 norm

Minmax Scaler

$$\frac{X - \cancel{X}_{min}}{X_{max} - X_{min}} \Rightarrow (0, 1)$$

Standard Scaler

$$\frac{X - \mu}{\sigma} \Rightarrow$$

mean is 0
variance is 1

Kmeans $\rightarrow$ substructure of the cluster
Geometry is not aligning with the actual boundary.

K modes

Mean value of features, we take mode

Distance = Hamming Distance

[Male, Youth, electronics]

1 + 0 + 1 = 2

[Female, Youth, furniture]

[female, Adult, furniture]

Centroid = [Female, Youth, Furniture]