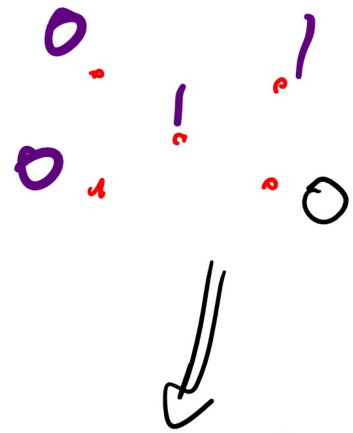
CS540 Introduction to Artificial Intelligence
Lecture 14

Young Wu
Based on lecture slides by Jerry Zhu and Yingyu Liang

July 2, 2019

Perceptron

Review



- Perceptron update rule.
- Perceptron termination condition.

iff data point linearly separable

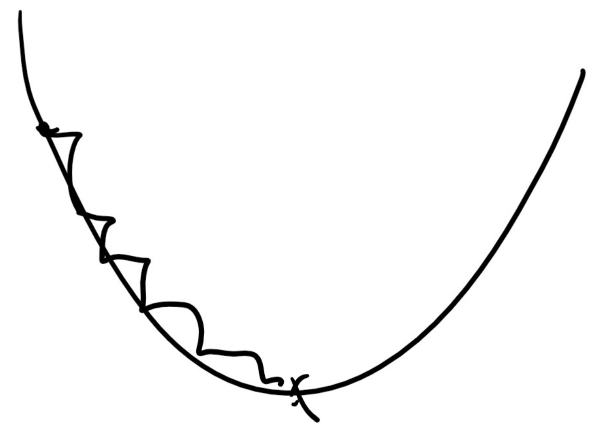
Logistic Regression

Review

- Logistic update rule.
- Logistic cost function.
- Convexity.
- Hessian, Laplacian, eigenvalue.

log cost.

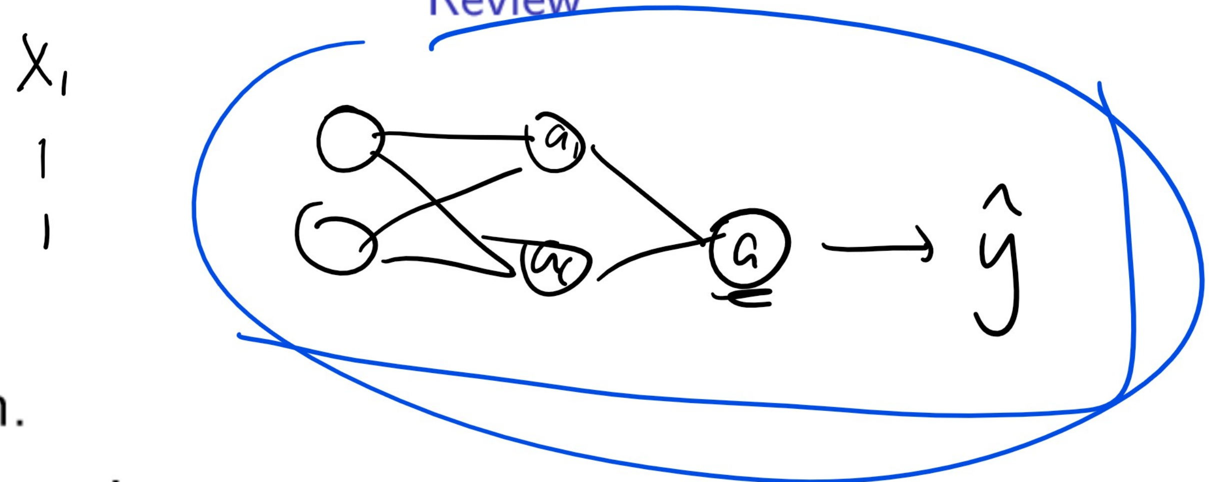
eigenvalue ≥ 0



$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

Neural Network

Review



- Activation.
- Backpropogation.
- L_1 and L_2 regularization.
- Cross validation.
- Multi class classification.

One vs One

one vs all

softmax

$$y = \begin{matrix}1\\2\\3\end{matrix} \Rightarrow \begin{pmatrix}1\\0\\0\end{pmatrix} \begin{pmatrix}0\\1\\0\end{pmatrix} \begin{pmatrix}0\\0\\1\end{pmatrix}$$

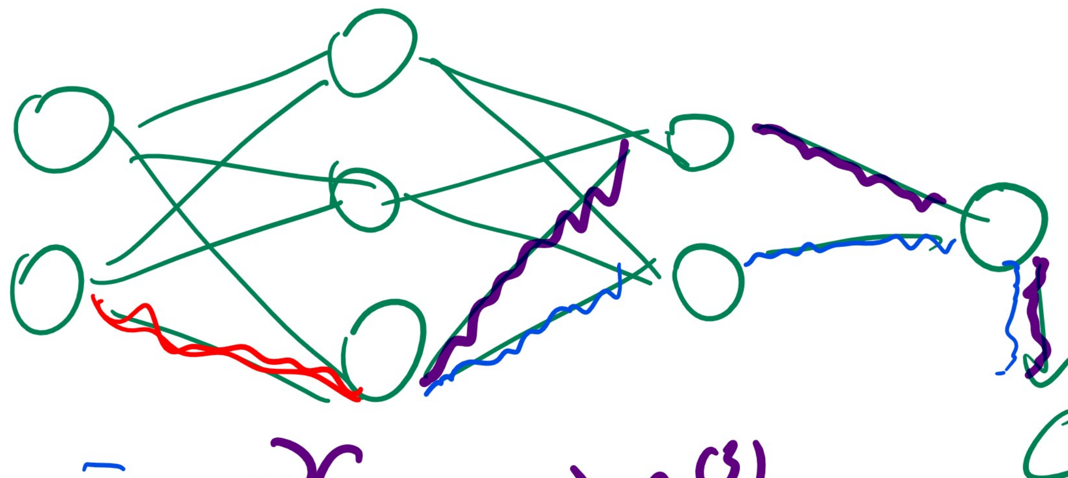
3 activation in last layer.

$$C = \sum_{i=1}^{n} \left(y_i - a_i^{(3)}\right)^2$$

Multi Layer Neural Network Example

Review

$a_1^{(2)} \cdot (1 - a_1^{(2)})\, w_{31}^{(2)}$



$$\boxed{a_1^{(2)} = g(\underbrace{w^T a^{(1)} + b}_{z_1^{(2)}})}$$

$$\frac{\partial C}{\partial w_{23}} = \frac{\partial C}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial a_3^{(1)}} \cdot \frac{\partial a_3^{(1)}}{\partial w_{23}}$$

$$+ \frac{\partial C}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial a_3^{(1)}} \cdot \frac{\partial a_3^{(1)}}{\partial w_{23}}$$

LTU Activation Example

Review

HW 2

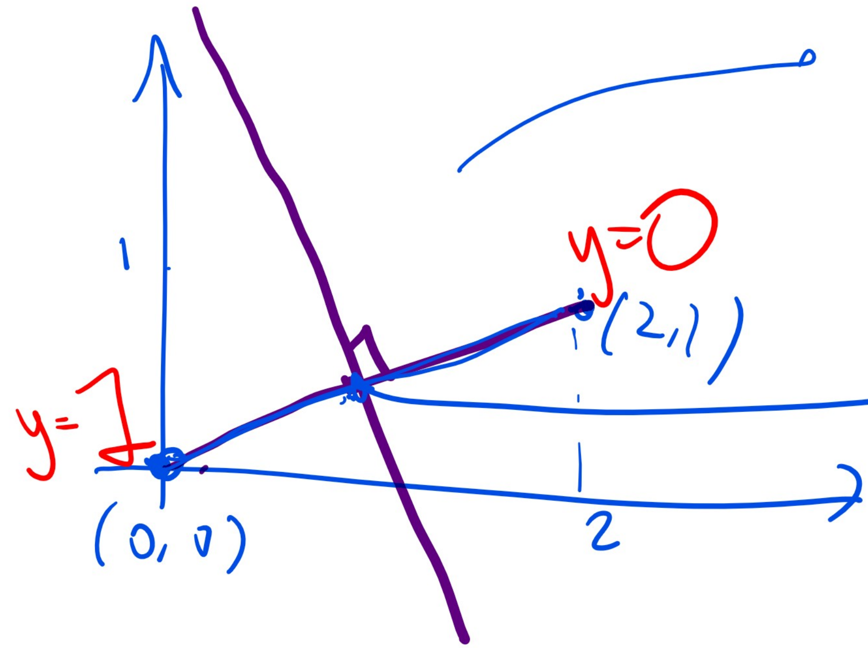
Support Vector Machine

Review

- Hard margin support vector.
- Soft margin maximization.
- Subgradient descent.
- Kernel trick.

Support Vector Margin Example

Review



slope $= -\frac{1}{\frac{1}{2}}$

$= -2$

point $= (1, \frac{1}{2})$

$x_2 = -2x_1 + b$

$\frac{1}{2} = -2 \cdot 1 + b$

$b = \frac{5}{2}$

SVM

$\mathbb{1}\{x_2 + 2x_1 - \frac{5}{2} \leq 0\}$

Feature Vector to Kernel Example

Review

$$\phi(x_1, x_2) = (x_1^2, \sqrt{2}\,x_1 x_2, x_2^2)$$

$$K = \phi^T(\overbrace{x_1, x_2}^{x})\,\phi(\overbrace{x_1', x_2'}^{x'}) = (x_1^2, \sqrt{2}\,x_1 x_2, x_2^2)\begin{pmatrix} x_1'^2 \\ \sqrt{2}\,x_1' x_2' \\ x_2'^2 \end{pmatrix}$$

$$= (x_1 x_1')^2 + 2(x_1 x_1')(x_2 x_2') + (x_2 x_2')^2$$

$$= (x_{i1} x_{i1}' + x_{i2} x_{i2}')^2$$

$$x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad x' = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$K(x, x') = 0$$

Decision Tree

Review

- Entropy.
- Information gain.
- Bagging and boosting.

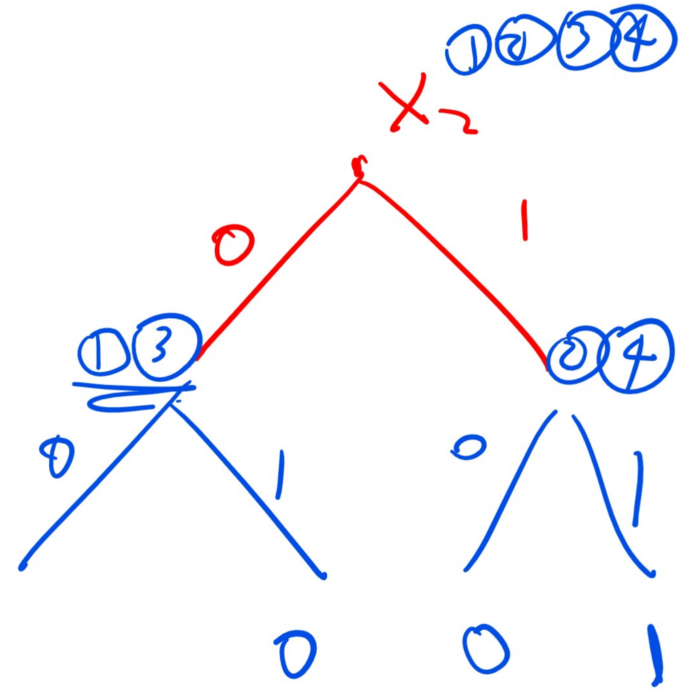
Decision Tree Example

Review

	x_1	x_2	y
1	0	0	1
2	0	1	0
3	1	0	0
4	1	1	1

~XOR

$\frac{1}{2}$ 1, $\frac{1}{2}$ 0



$H(Y)$

$$H(Y \mid X_2) = \Pr\{X_2 = 0\} \cdot \underline{H(Y \mid X_2 = 0)} + \Pr\{X_2 = 1\} \cdot H(Y \mid X_2 = 1)$$

$$= \frac{1}{2}\Big(\underbrace{-\frac{1}{2}\log\frac{1}{2}}_{Y=0} \underbrace{- \frac{1}{2}\log\frac{1}{2}}_{Y=1}\Big) + \frac{1}{2}(1)$$

$$\left(\frac{1}{2} + \frac{1}{2}\right) = 1$$

$$= 1$$

K Nearest Neighbor

Review

L_1, L_2, L_∞

- Distance functions.

K Nearest Neighbor Cross Validation Example

Review

1-nn

3-fold CV

$$\begin{array}{c|cc|cc|cc} & \multicolumn{2}{c|}{S_1} & \multicolumn{2}{c|}{S_2} & \multicolumn{2}{c}{S_3} \\ x_i = & 1 & 2 & 3 & 4 & 5 & 6 \\ y_i = & \cancel{0} & \cancel{0} & \cancel{1} & 1 & 1 & \cancel{0} \end{array}$$

3.5

1

1/2

train on S_2, S_3, test on S_1 $\Rightarrow$

S_2

S_3

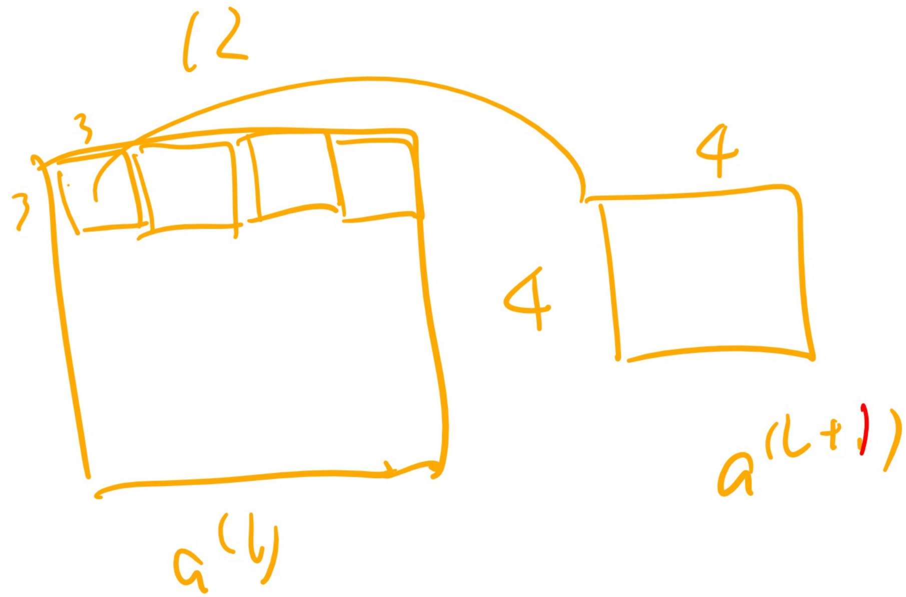
$$\left.\right\} \text{accuracy} = \frac{2}{6} = \frac{2}{3}.$$

+1/2

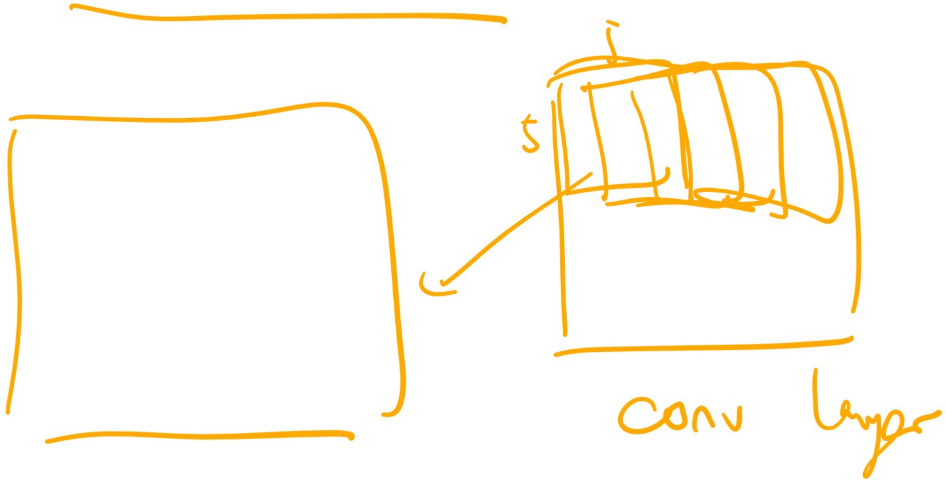
+1

Convolutional Neural Network

Review



- Convolution.
- Pooling.
- Trained weights.



trained weights in conv layer

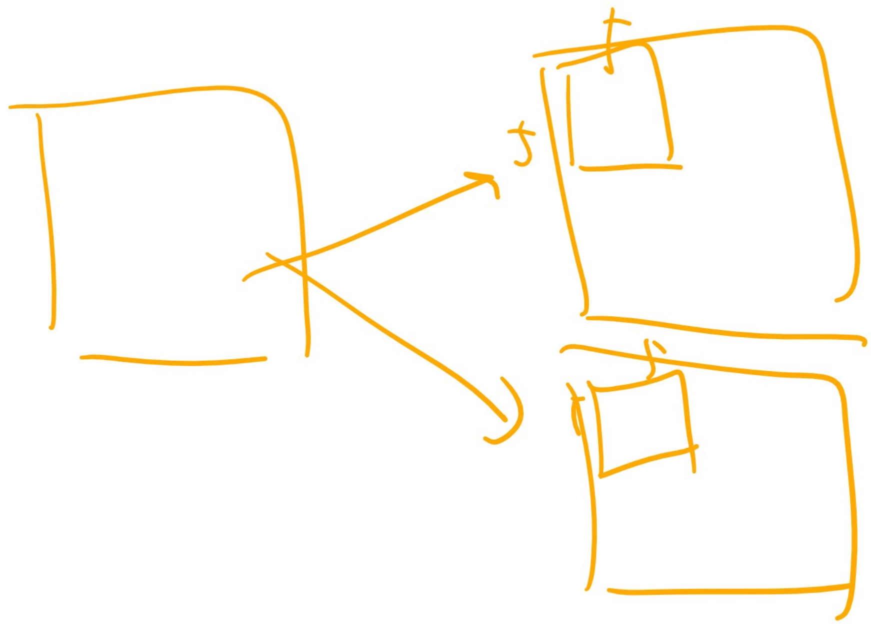
= # elements of filters

= 25

weights in pooling = 0

Convolutional Weights Count Example

Review

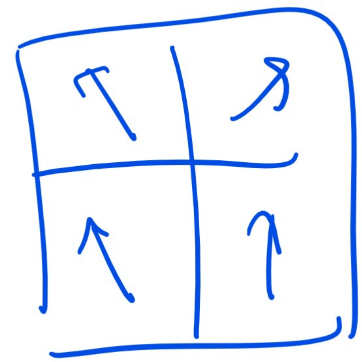


$5 \times 5 \times 2 = 50$

Computer Vision

Review

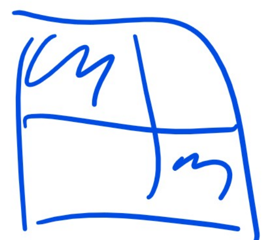
HW4



- Histogram of Gradients Features.
- Scale Invariant Feature Transform.
- ~~Block normalization.~~
- Dominant orientation.
- Harr Features.

add up G

not counting



Histogram of Gradient Example

Review

Natural Language Processing

Review

HW5

- Bigram and trigram model.
- Transition matrix.
- Random word generation.
- Bayes rule.

$P_s\{a \mid b\}$

$$\frac{C_{ba} + 1}{C_b + |\text{vocabulary}|}$$

a b b b c $\nearrow$ 3

Document Bayes Rule Example

Review

A	70%	"the"	$\frac{1}{3}$
B	20%	"the"	$\frac{1}{3}$
C	10%	"the"	$\frac{1}{3}$

$$\Pr\{B \mid \text{"the"}\} = \frac{\Pr\{B, \text{"the"}\}}{\Pr\{\text{"the"}\}}$$

$\Pr\{B\} \cdot \Pr\{\text{"the"} \mid B\}$

$= \Pr\{\text{"the"}, A\} + \Pr\{\text{"the"}, B\} + \Pr\{\text{"the"}, C\}$

$$= \boxed{\frac{\frac{1}{3} \cdot 0.2}{\frac{1}{3} \cdot 0.7 + \frac{1}{3} \cdot 0.2 + \frac{1}{3} \cdot 0.1}}$$

Bayesian Network

Review

$$\frac{C_{AB}}{C_B}$$

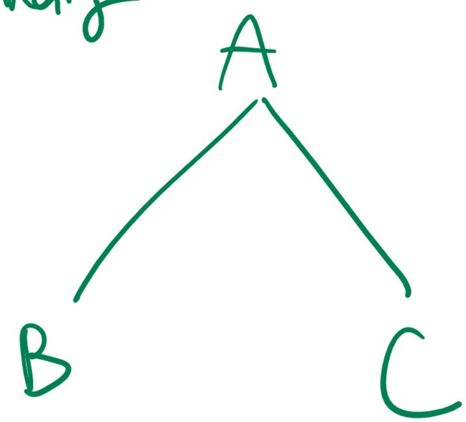
- Conditional probability table.
- Maximum likelihood estimation.
- Training vs inference.
- Chow Liu algorithm.

max Spanng tree.

Common Cause Network Example

Review

A, B, C binary
0, 1



CPT

$$\Pr\{A\} = 0.1$$

$$\Pr\{B \mid A\} = 0.2, \quad \Pr\{\neg B \mid A\} = 0.8$$

$$\Pr\{B \mid \neg A\} = 0.3$$

$$\Pr\{C \mid A\} = 0.4$$

$$\Pr\{C \mid \neg A\} = 0.5$$

$$\Pr\{B \mid \neg C\} = \frac{\Pr\{B, \neg C, A = 0,1\}}{\Pr\{\neg C, A = 0,1\}}$$

$$\Pr\{B, \neg C, A = 0,1\} = \Pr\{B, \neg C, A\} + \Pr\{B, \neg C, \neg A\}$$

$$= \Pr\{B \mid A\} \cdot \Pr\{\neg C \mid A\} \cdot \Pr\{A\} + \Pr\{B \mid \neg A\} \cdot \Pr\{\neg C \mid \neg A\} \cdot \Pr\{\neg A\}$$

$$0.2 \cdot 1 - 0.4 \cdot 0.1 + 0.3 \cdot (1 - 0.5) \cdot (1 - 0.1)$$

$$\Pr\{\neg C, A\} + \Pr\{\neg C, \neg A\}$$

$$= (1-0.4)\cdot 0.1 + (1-0.5)\cdot(1-0.1)$$