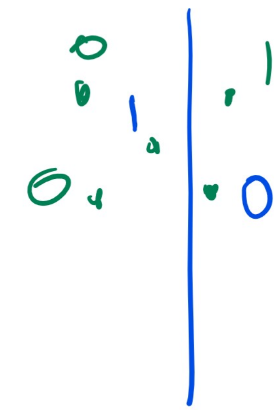
CS540 Introduction to Artificial Intelligence
Lecture 14

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

July 2, 2019

Perceptron

Review



- Perceptron update rule.
- Perceptron termination condition. ← iff linearly separable.

Logistic Regression

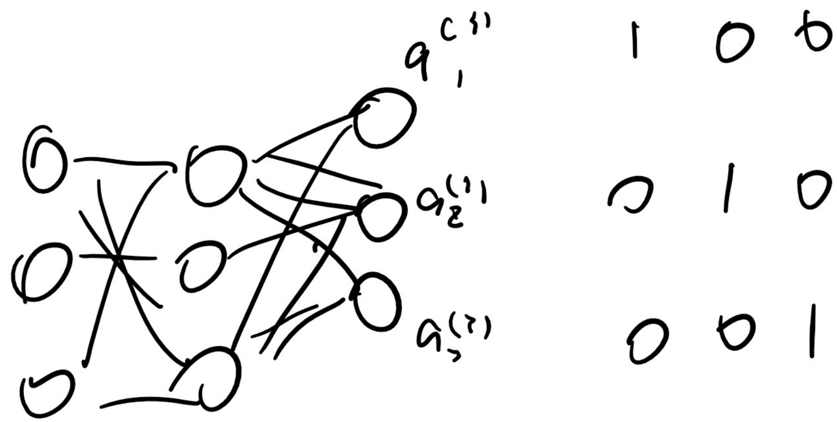
Review

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

Neural Network

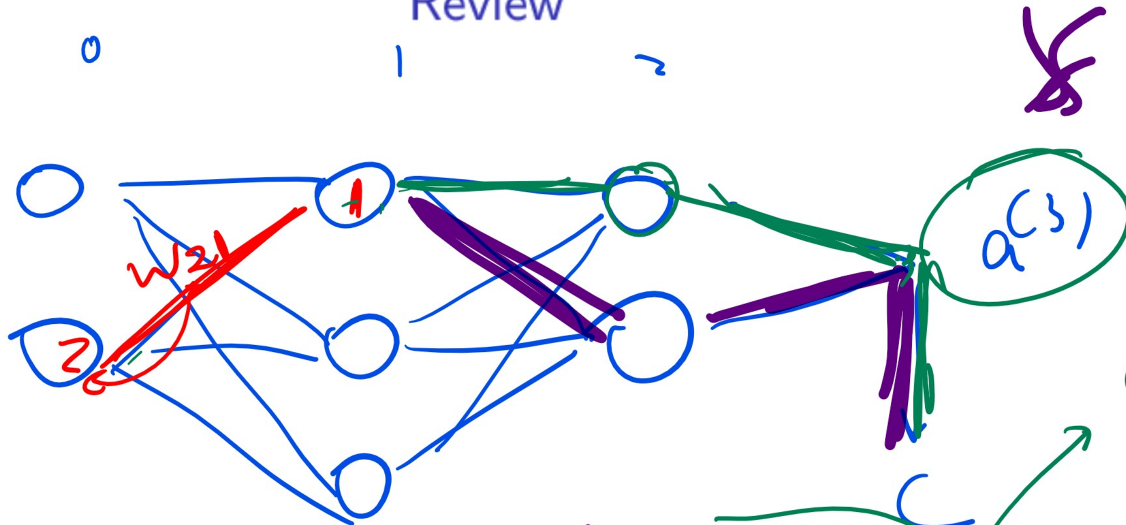
Review

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



Multi Layer Neural Network Example

Review



$$a_2^{(2)} = g(z_2^{(2)})$$

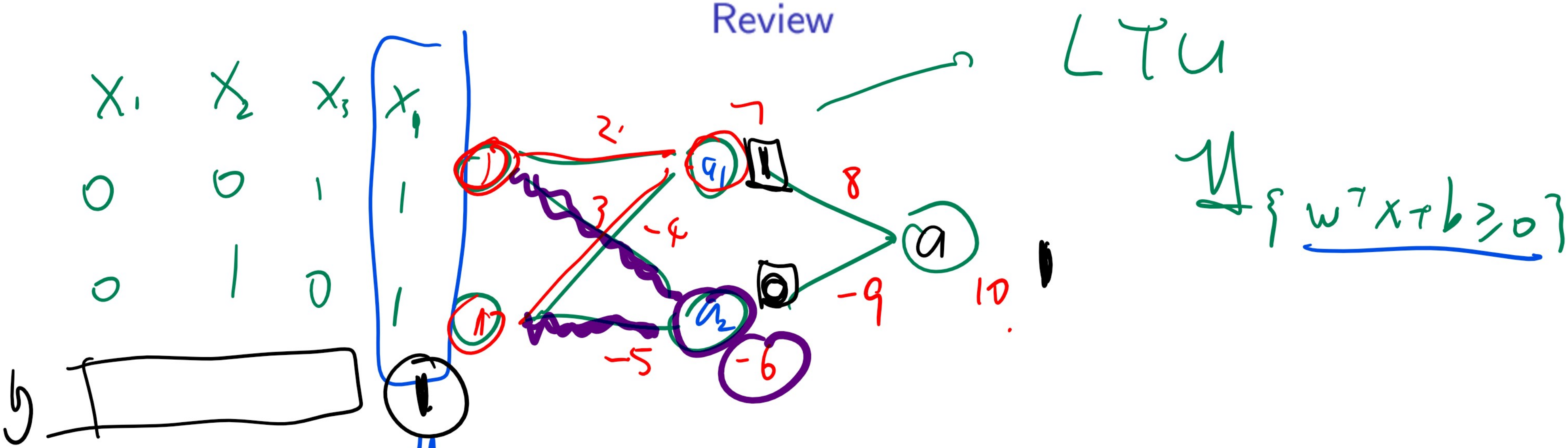
$$z_2^{(2)} = w_{12}^{(2)} a_1^{(1)} + \dots$$

$$\frac{\partial C}{\partial w_{21}} = \frac{\partial C}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial a_2^{(2)}} \cdot \frac{\partial a_2^{(2)}}{\partial a_1^{(1)}} \cdot \frac{\partial a_1^{(1)}}{\partial w_{21}}$$

$$+ \frac{\partial C}{\partial a^{(3)}} \cdot \frac{\partial a^{(3)}}{\partial a_1^{(2)}} \cdot \frac{\partial a_1^{(2)}}{\partial a_1^{(1)}} \cdot \frac{\partial a_1^{(1)}}{\partial w_{21}}$$

LTU Activation Example

Review



$$a_1 = \mathbb{1}\{1\cdot 2 + 1\cdot 3 + 7 \geq 0\} = 1$$

$$a_2 = \mathbb{1}\{1\cdot -4 + 1\cdot -5 - 6 \geq 0\} = 0$$

False

$$a = \mathbb{1}\{\underbrace{1\cdot 8 + 0\cdot(-9) + 10}_{\geq 0} \geq 0\} = 1$$

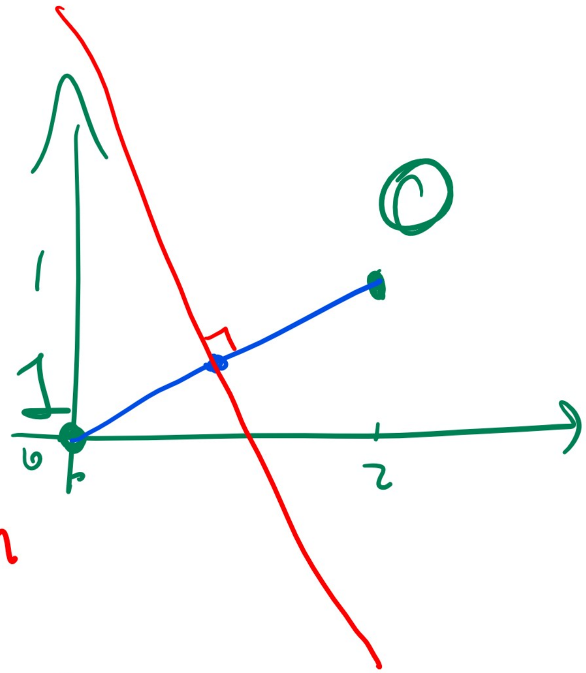
Support Vector Machine

Review

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

Support Vector Margin Example

Review



slope $\frac{1}{2}$

SVM slope -2

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

$$x_2 = -2x_1 + b$$

$$\frac{1}{2} = -2 \cdot 1 + b \Rightarrow b = \frac{5}{2}$$

SVM

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

$$= \mathbb{1}\{\frac{4}{5}x_1 + \frac{2}{5}x_2 - 1 \le 0\}$$

$$= \mathbb{1}\{-\frac{4}{5}x_1 - \frac{2}{5}x_2 + 1 \ge 0\}$$

Feature Vector to Kernel Example

Review

~~Kernel~~

$$\phi(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

Kernel

$$\phi^T(x_1, x_2)\,\phi(x_1', x_2') = (x_1^2, \sqrt{2}x_1x_2, x_2^2)\begin{pmatrix} x_1'^2 \\ \sqrt{2}x_1'x_2' \\ x_2'^2 \end{pmatrix}$$

instance 1 instance 2

$$= (x_1x_1')^2 + 2(x_1x_1'x_2x_2') + (x_2x_2')^2$$

$$= (x_1x_1' + x_2x_2')^2$$

instances

$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$$K = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

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

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

Decision Tree

Review

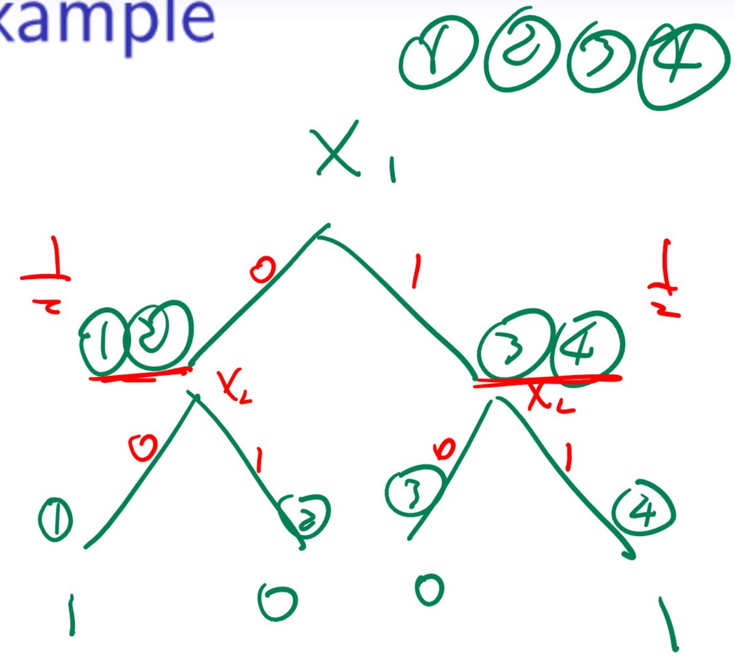
- Entropy.
- Information gain.
- Bagging and boosting.

Decision Tree Example

Review

NXOR

	X_1	X_2	Y
①	0	0	1
②	0	1	0
③	1	0	0
④	1	1	1



$H(Y \mid X_1) = 1$

$= \frac{1}{2} H(Y \mid X_1 = 0) + \frac{1}{2} H(Y \mid X_1 = 1)$

entropy of ①② — entropy of ③④

$-\left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)$

$+1$ — $+1$

$H(Y) = 1$

$IG = 0$

K Nearest Neighbor

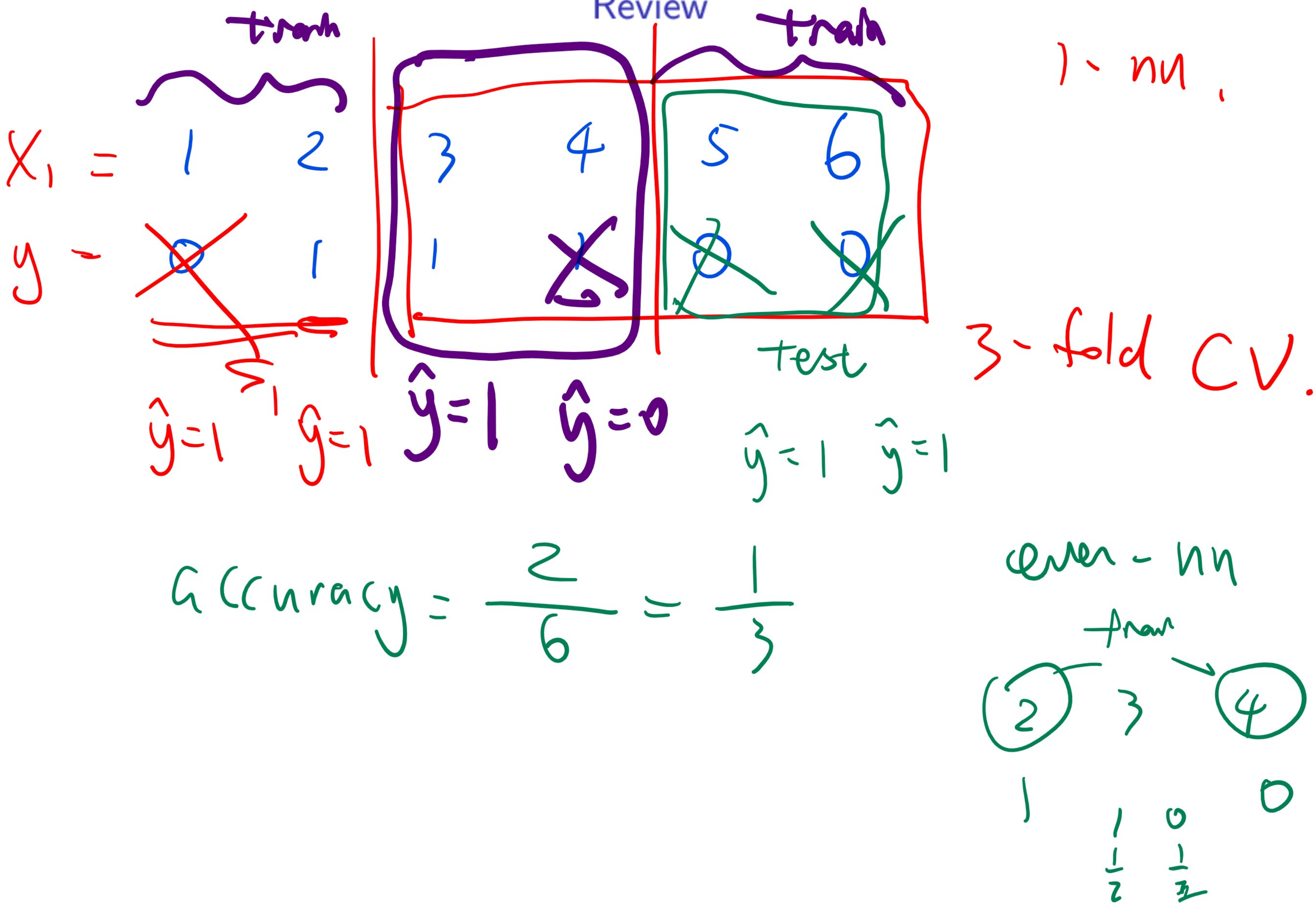
Review

- Distance functions.

  L_1, L_2, L_∞

K Nearest Neighbor Cross Validation Example

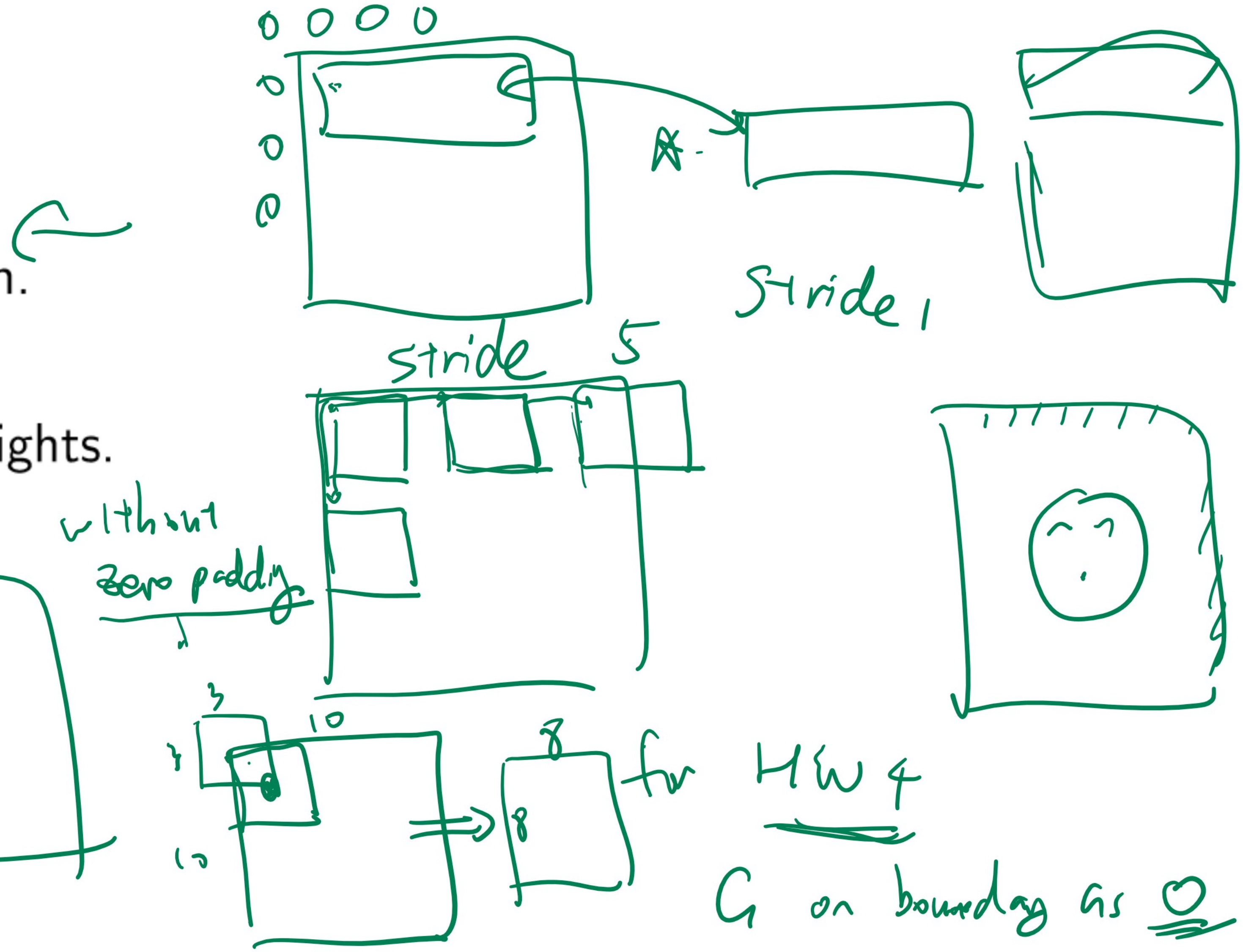
Review

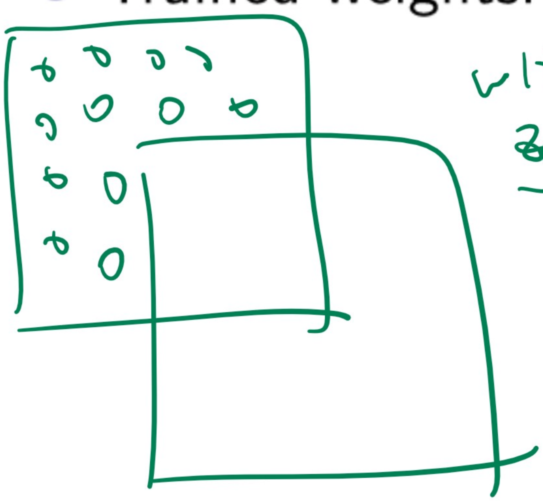


Convolutional Neural Network

Review

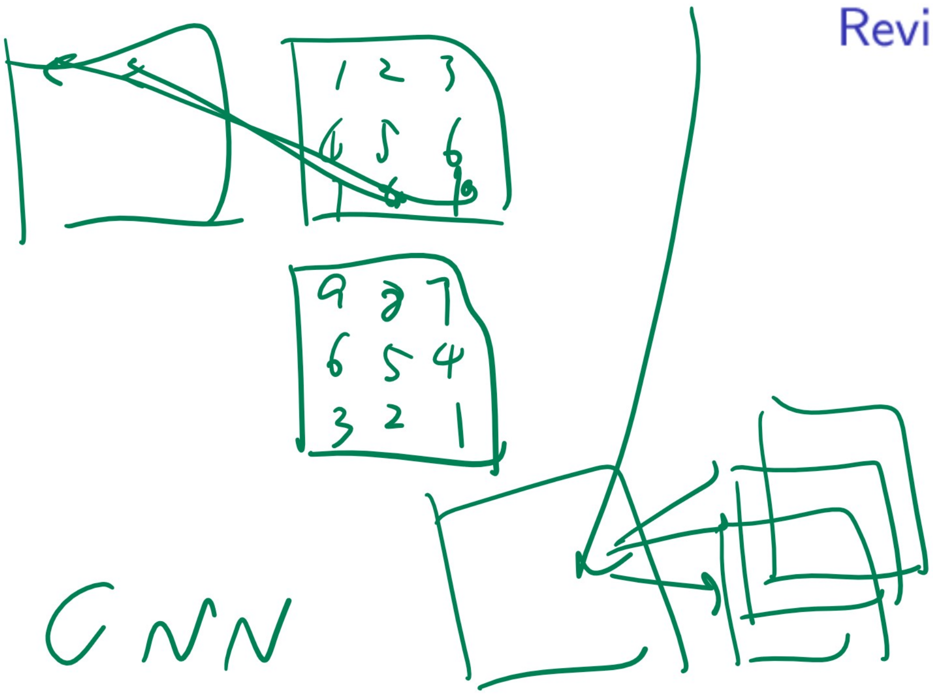
- Convolution.
- Pooling.
- Trained weights.

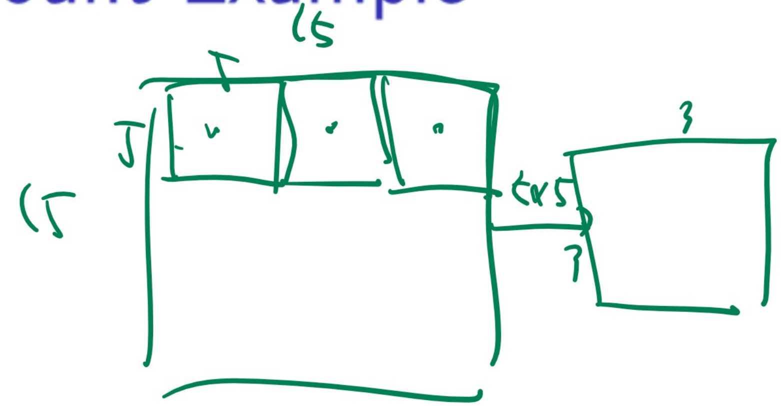




Convolutional Weights Count Example

Review





15×15 activation map
matrix

5×5 pooling filter

3×3 activation map.

CNN

↓

nonoverlapping → stride 5

filters conv layer all need trained

pool none

5×5 filter

train 25 weights.

fully connected all need trained

Computer Vision

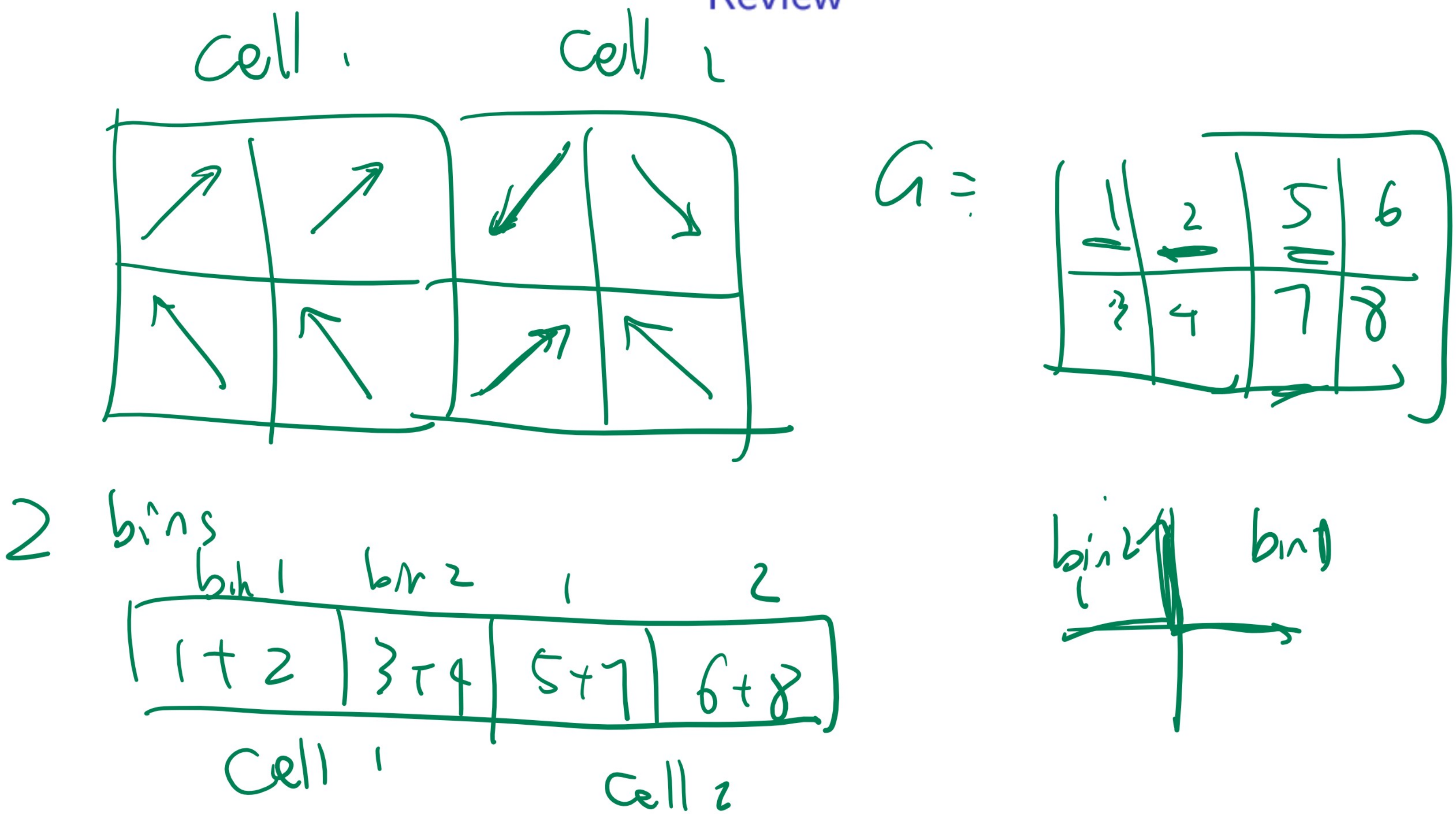
Review

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

gee histogram

Histogram of Gradient Example

Review



Natural Language Processing

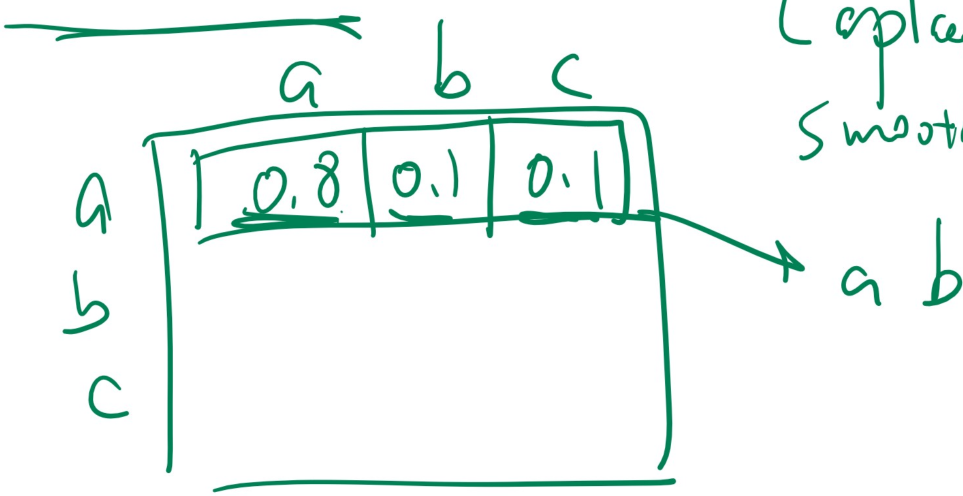
Review

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

$$\Pr\{w \mid a\} = \frac{C_{aw}}{C_a}$$

Laplace smooth

$$\frac{C_{aw} + 1}{C_a + |\text{vocabulary}|}$$



Document Bayes Rule Example

Review

$$\frac{1}{3}\left[\begin{array}{lll} A & \rightarrow & 0.7 \; H \\ B & \rightarrow & 0.5 \; H \\ C & \rightarrow & 0.2 \; H \end{array}\right.$$

$$\Pr\{B \mid H\} = \frac{\Pr\{BH\}}{\Pr\{H\}}$$

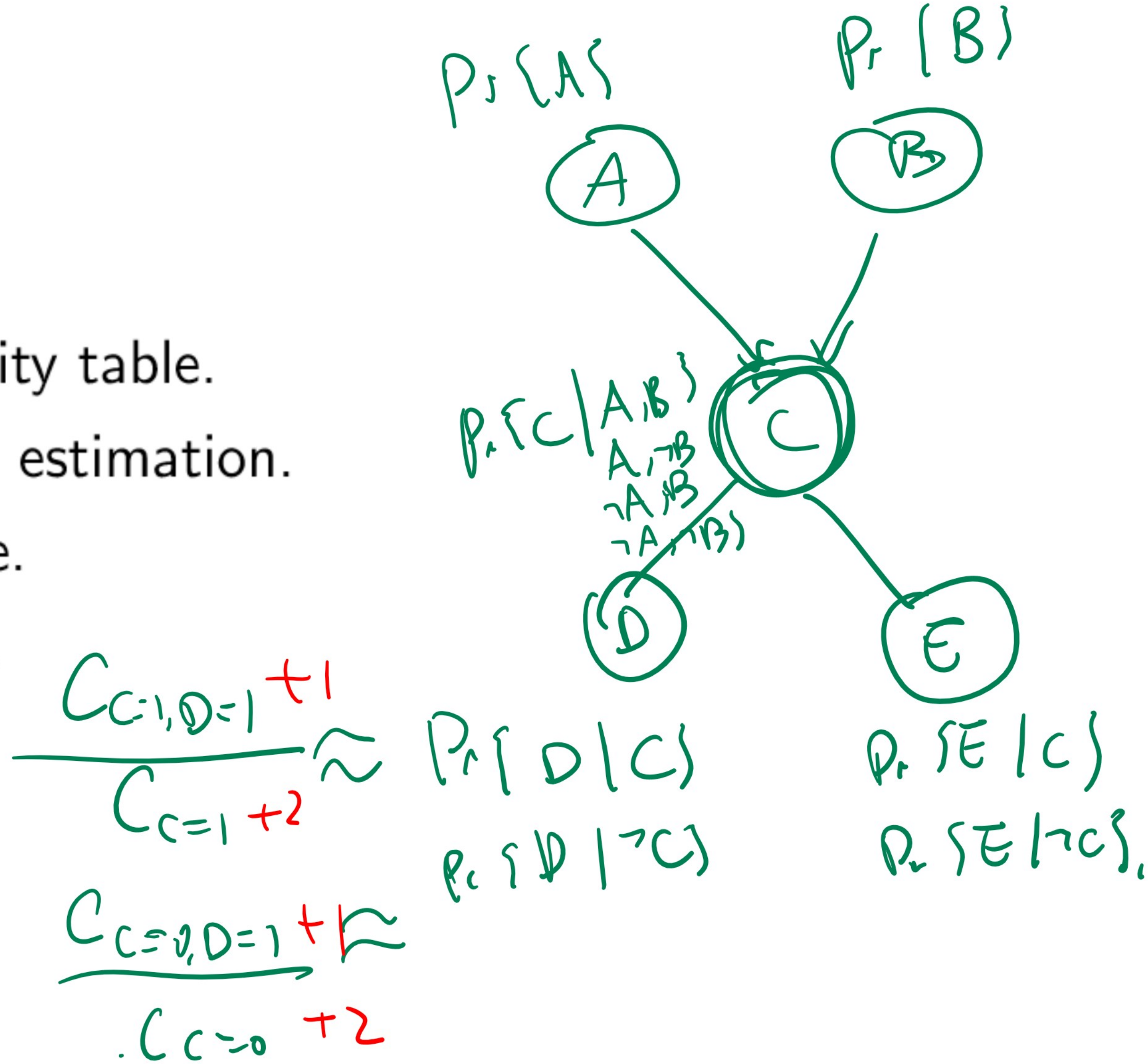
$$\Pr\{BH\} \Rightarrow \underbrace{\Pr\{H \mid B\}}_{\frac{1}{2}} \; \underbrace{\Pr\{B\}}_{\frac{1}{3}}$$

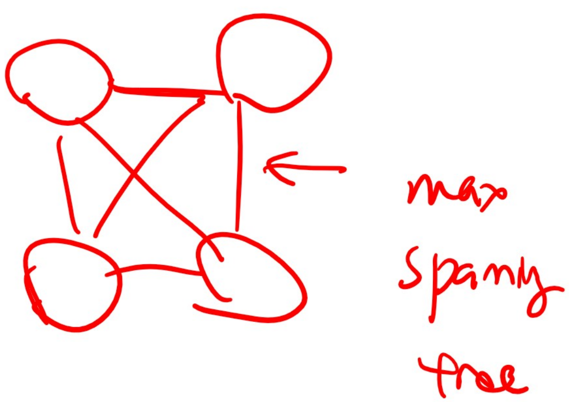
$$\Pr\{H\} \Rightarrow \frac{1}{3}\cdot 0.7 + \frac{1}{3}\cdot 0.5 + \frac{1}{3}\cdot 0.2$$

Bayesian Network

Review

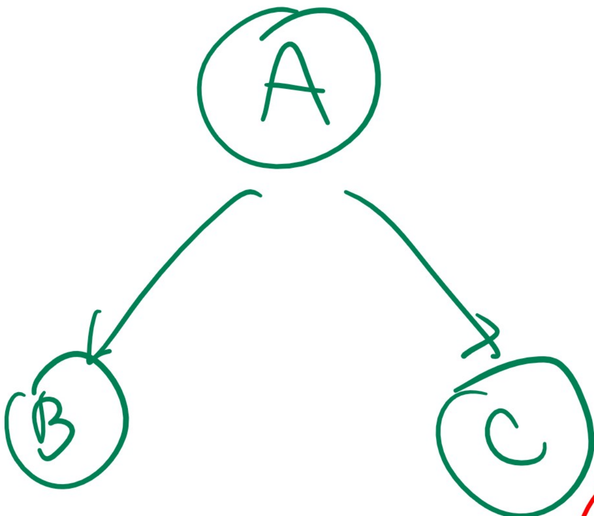
- Conditional probability table.
- Maximum likelihood estimation.
- Training vs inference.
- Chow Liu algorithm.





Common Cause Network Example

Review



$\Pr\{A\} = 0.7$

$\Pr\{B \mid A\} = 0.3 \qquad \Pr\{C \mid A\} = 0.4$

$\Pr\{B \mid \neg A\} = 0.2 \qquad \Pr\{C \mid \neg A\} = 0.6$

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

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

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

$$\Pr\{B, \neg C\} = 0.3 \quad 0.6 \quad 0.7 + 0.2 \quad 0.4 \quad 0.3$$

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

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

$$\Pr\{\neg C\} = 0.6 \cdot 0.7 + 0.4 \cdot 0.3$$