Lecture 23: Exam 3 Review

Mark Hasegawa-Johnson

University of Illinois

ECE 417: Multimedia Signal Processing

Outline

1. Administrative
2. Topics covered
3. Sample Problems

Wnen and Where?

- Friday, December 8
- 1:30-4:30pm
- Here (ECEB 2013)

- Bring:
  - Up to 3 sheets of notes, hand-written or 12pt+ notes on both sides
  - Pencils or pens
- Don't bring:
  - Calculators, computers, tablets, phones

Is it comprehensive?

Yes, but with an emphasis on the last third of the course.

- Total: 200 points
- About 34 points: First third of the course
- About 34 points: Second third of the course
- About 132 points: Last third of the course

Outline

What is the "last third"?

- Covered on the exam:
  - Lecture 15: PCA
  - Lecture 17: RNN
  - Lecture 18: LSTM
- Not covered:
  - Lecture 19: Speaker verification
  - Lecture 20: AutoVC
  - Lecture 21: Transformer
  - Lecture 22: Self-supervised learning

PCA

- Symmetric positive semidefinite matrices:

$$\Sigma = \boldsymbol{U}\Lambda\boldsymbol{U}^T, \quad \boldsymbol{U}^T\Sigma\boldsymbol{U} = \Lambda, \quad \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}$$

- Centered dataset:

$$\boldsymbol{X} = [\boldsymbol{x}_1 - \boldsymbol{\mu}, \ldots, \boldsymbol{x}_M - \boldsymbol{\mu}], \quad \Sigma = \frac{1}{M-1}\boldsymbol{X}\boldsymbol{X}^T, \quad \boldsymbol{G} = \boldsymbol{X}^T\boldsymbol{X}$$

- Singular value decomposition:

$$\boldsymbol{X} = \boldsymbol{U}\Lambda^{1/2}\boldsymbol{V}^T$$

- The principal components are the first K elements of $\boldsymbol{y}_m = \boldsymbol{U}^T(\boldsymbol{x}_m - \boldsymbol{\mu})$. The amount of energy they capture is:

$$\frac{1}{M-1}\sum_{m=1}^{M}\|\boldsymbol{y}_m\|^2 = \sum_{k=1}^{K}\lambda_k$$

RNN

- Back-Prop, in general, is just the chain rule of calculus:

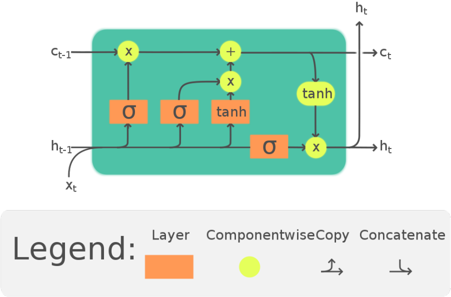
$$\frac{d\mathcal{L}}{dw} = \sum_{i=0}^{N-1} \frac{d\mathcal{L}}{dh_i}\frac{\partial h_i}{\partial w}$$

- Convolutional Neural Networks are the nonlinear version of an FIR filter. Coefficients are shared across time steps.
- Recurrent Neural Networks are the nonlinear version of an IIR filter. Coefficients are shared across time steps. Error is back-propagated from every output time step to every input time step.

$$\frac{d\mathcal{L}}{dh[n]} = \frac{\partial\mathcal{L}}{\partial h[n]} + \sum_{m=1}^{M} \frac{d\mathcal{L}}{dh[n+m]}\frac{\partial h[n+m]}{\partial h[n]}$$

$$\frac{\partial\mathcal{L}}{\partial w[m]}(w[1],\ldots,w[M]) = \sum_{n} \frac{d\mathcal{L}}{dh[n]}\frac{\partial h[n]}{\partial w[m]}$$

Neural Network Model: LSTM



$$i[t] = \text{input gate} = \sigma(w_i x[t] + u_i h[t-1] + b_i)$$
$$o[t] = \text{output gate} = \sigma(w_o x[t] + u_o h[t-1] + b_o)$$
$$f[t] = \text{forget gate} = \sigma(w_f x[t] + u_f h[t-1] + b_f)$$
$$c[t] = f[t]c[t-1] + i[t]\tanh\left(w_c x[t] + u_c h[t-1] + b_c\right)$$
$$h[t] = \text{output} = o[t]\tanh(c[t])$$

Outline

1 Administrative

2 Topics covered

3 Sample Problems

Sample Problems

- Sample problems about PCA
- Sample problems about RNN
- Sample problems about LSTM