CS221 Section 1

Foundations

Roadmap

Recurrence Relation

Continuous Optimization

Probability Theory

Python

Coin Payment

Problem

Suppose you have an unlimited supply of coins with values 2, 3, and 5 cents

How many ways can you pay for an item costing 12 cents?

Coin Payment

What if the order ...

... **matters**?

... **does not matter**?

Recurrence Relation: Break down into smaller problems

Memoization: Remember what you already calculated

Roadmap

Recurrence Relation

Continuous Optimization

Probability Theory

Python

Multiclass Classification

Let's review binary classification

$$x \longrightarrow \boxed{f_{\mathbf{w}}} \longrightarrow y \in \{-1, +1\}$$

Score:

$$\text{score}_{+1}(x, \mathbf{w}) = \mathbf{w} \cdot \phi(x)$$
$$\text{score}_{-1}(x, \mathbf{w}) = (-\mathbf{w}) \cdot \phi(x)$$

Prediction:

$$f_{\mathbf{w}}(x) = \begin{cases} +1 & \text{if score}_{+1}(x, \mathbf{w}) > \text{score}_{-1}(x, \mathbf{w}) \\ -1 & \text{otherwise} \end{cases}$$

$$f_{\mathbf{w}}(x) = \arg \max_{y \in \{-1,+1\}} \text{score}_y(x, \mathbf{w})$$

Multiclass Classification

Problem

Suppose we have 3 possible labels $y \in \{\mathsf{R}, \mathsf{G}, \mathsf{B}\}$

Weight vectors: $\mathbf{w} = \{\mathbf{w}_{\mathsf{R}}, \mathbf{w}_{\mathsf{G}}, \mathbf{w}_{\mathsf{B}}\}$

Scores: $[\mathbf{w}_{\mathsf{R}} \cdot \phi(x)], [\mathbf{w}_{\mathsf{G}} \cdot \phi(x)], [\mathbf{w}_{\mathsf{B}} \cdot \phi(x)]$

Prediction: $\hat{y} = f_{\mathbf{w}}(x) = \arg \max_{y \in \{\mathsf{R}, \mathsf{G}, \mathsf{B}\}} [\mathbf{w}_y \cdot \phi(x)]$

Multiclass Classification

How to learn $\mathbf{w}$?

How about **0-1 loss**:

$$\text{Loss}_{\text{0-1}}(x, y, \mathbf{w}) = \begin{cases} 1 & \text{if } \hat{y} \neq y \\ 0 & \text{otherwise} \end{cases}$$

Problem: Gradient is 0 almost everywhere

Multiclass Classification

How to learn $\mathbf{w}$?

Recall **hinge loss**:

$$\text{margin} = \text{score}_y(x, \mathbf{w}) - \max_{y' \neq y} \text{score}_{y'}(x, \mathbf{w})$$

$$\text{Loss}_{\text{Hinge}}(x, y, \mathbf{w}) = \max\{1 - \text{margin}, 0\}$$

What is the gradient? Let $y^* = \arg\max_{y' \neq y} \text{score}_{y'}(x, \mathbf{w})$

$$\nabla_{\mathbf{w}_y} \text{Loss}_{\text{Hinge}}(x, y, \mathbf{w}) = \begin{cases} -\phi(x) & 1 - \text{margin} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\nabla_{\mathbf{w}_{y^*}} \text{Loss}_{\text{Hinge}}(x, y, \mathbf{w}) = \begin{cases} +\phi(x) & 1 - \text{margin} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{other } y' : \nabla_{\mathbf{w}_{y'}} \text{Loss}_{\text{Hinge}}(x, y, \mathbf{w}) = 0$$

Roadmap

Recurrence Relation

Continuous Optimization

Probability Theory

Python

Random Variables

Discrete:

$$\mathbb{P}(A = a) \qquad \text{or} \qquad p_A(a)$$

Continuous:

$$\xcancel{\mathbb{P}(A = a)}$$

$$f_A(a)$$

$$\mathbb{P}(A \leq c) = \int_{--}^{c} f_A(a)\, da$$

Random Variables

	A = 0	**A = 1**	**A = 2**	**A = 3**
B = 0	0.1	0.25	0.1	0.05
B = 1	0.15	0	0.15	0.2

- What is $\mathbb{P}(A = 2)$
- What is $\mathbb{P}(A = 2 \mid B = 1)$

Random Variables

Independence:

$$\forall a, b, \quad \mathbb{P}(A = a, B = b) = \mathbb{P}(A = a)\mathbb{P}(B = b)$$

$$\forall a, b, \quad f_{A,B}(a, b) = f_A(a) f_B(b)$$

Expectation:

$$\mathbb{E}[A] = \sum_a a\,\mathbb{P}[A = a]$$

$$\mathbb{E}[A] = \int a f_A(a)\, da$$

Random Variables

	A = 0	**A = 1**	**A = 2**	**A = 3**
B = 0	0.1	0.25	0.1	0.05
B = 1	0.15	0	0.15	0.2

- Are A and B independent?
- What are $\mathbb{E}[A]$, $\mathbb{E}[B]$, $\mathbb{E}[A+B]$

Linearity of Expectation: $\mathbb{E}[A+B] = \mathbb{E}[A] + \mathbb{E}[B]$

True even when A and B are dependent!

Random Variables

Variance:

$$\mathrm{Var}[A] = \mathbb{E}[(A - \mathbb{E}[A])^2] = \mathbb{E}[A^2] - \mathbb{E}[A]^2$$

Covariance:

$$\begin{aligned}\mathrm{Cov}[A, B] &= \mathbb{E}[(A - \mathbb{E}[A])(B - \mathbb{E}[B])] \\ &= \mathbb{E}[AB] - \mathbb{E}[A]\mathbb{E}[B]\end{aligned}$$

If $\mathrm{Cov}[A, B] = 0$, we say A and B are **uncorrelated**

Random Variables

If A and B are independent, then

- $\mathrm{Cov}[A, B] = \mathbb{E}[AB] - \mathbb{E}[A]\mathbb{E}[B] = 0$

  Independence implies uncorrelatedness

- $\mathrm{Var}[A + B] = \mathrm{Var}[A] + \mathrm{Var}[B]$

  Noise adds up

But the converse is **not** true!

Roadmap

Recurrence Relation

Continuous Optimization

Probability Theory

Python

Syntactic Sugar

- List comprehension
- List slicing
- Passing functions
- Reading and writing files

Gotchas

- Integer division
- Tied objects
- Global variables

References

- Official Documentation (has a tutorial):

  https://docs.python.org/2.7/

- Learn X in Y minutes:

  http://learnxinyminutes.com/docs/python/

- You don't need to know `numpy`. But if you want to:

  http://nbviewer.ipython.org/gist/rpmuller/5920182

Matrix Calculus

$$f(\mathbf{w}) = (\mathbf{a} \cdot \mathbf{w} + 1)^2 + b \|\mathbf{w}\|_2^2 + \mathbf{w}^\top C \mathbf{w}$$

Compute $\nabla_{\mathbf{w}} f(\mathbf{w})$

$$\boxed{\begin{aligned} \nabla_{\mathbf{w}} \mathbf{a} \cdot \mathbf{w} &= \mathbf{a} \\ \nabla_{\mathbf{w}} \|\mathbf{w}\|_2^2 = \nabla_{\mathbf{w}} \mathbf{w} \cdot \mathbf{w} &= 2\mathbf{w} \\ \nabla_{\mathbf{w}} \mathbf{w}^\top C \mathbf{w} &= (C + C^\top)\mathbf{w} \end{aligned}}$$

Questions?