

Recommender Systems Mathematical Foundations

Professor Robin Burke Spring 2019

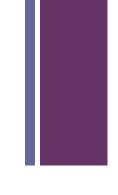
+ Outline

- Derivatives
- **■** Gradients
- Gradient descent example

+ Derivative

- A generalization of the idea of slope
 - A line has a slope
 - https://en.wikipedia.org/wiki/Slope#/media/File:Gradient of a line in coordinates from -12x%2B2 to %2B12x%2B2.gif
- Does a curve has slope
 - Yes, but not "a slope"
 - Different slopes at different points
 - https://en.wikipedia.org/wiki/Slope#/media/File:Tangent_function_animation.gif
 - "Instantaneous slope"

Derivative



- A function f' that yields the slope of the curve f at each point
- Examples
 - f = 5x
 - **■** f' = ?
 - f = 5x + 10
 - f' = ?
 - f = -5x 10
 - f' = ?
 - f = 5x + 6y
 - **■** f' = ?
 - Other notation
 - $\blacksquare \frac{df}{dx}$
 - Make it possible to think about functions of multiple variables

Derivative

$$f' = 6x$$

- Classic proof
 - slope = rise / run
 - slope = $f(x+\Delta) f(x) / \Delta$

$$f'(x) = \frac{(3x^2 + 6x\Delta + 3\Delta^2) - 3x^2}{\Delta} = \frac{6x\Delta + 3\Delta^2}{\Delta} = 6x + 3\Delta$$

■ Now we let $\Delta \rightarrow 0$

General exponent rule



$$f(x) = ax^n$$

$$f'(x) = anx^{n-1}$$

- Lots of other rules for computing derivatives
 - Take a calculus class!



Derivative of 5x⁻³



- B: 15x⁻²
- **C**: 15x⁻⁴
- D: -15x⁻⁴

Slope at a point



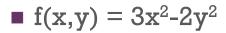
- Slope of $3x^2$ where x = 7?
 - f'(7) = 6(7) = 42
- Important question
 - when is the slope = 0
 - \bullet 6x = 0,
 - only when x = 0

For the purposes of this class, why is slope = 0 important?

- A: Because that's the coolest point
- B: Because that's the maximum or minimum of the function
- C: Because that's where the function becomes undefined
- D: Because that's the maximum or minimum of the function if it is convex



+ Derivatives



■ More technically correct Partial Derivatives



Gradient



$$y = f(x_1, x_2, x_3, x_4) = \sum_{i=1}^4 x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

- A: 2x
- \blacksquare B: $2x_1$
- **C**: 2x_i
- \blacksquare D: $2x_1 + x_2^2 + x_3^2 + x_4^2$

Gradient



- Separate function for each dimension
- We can evaluate at a given point
 - $\nabla f(1,1,1,2) = [2,2,2,4]$



Our big hairy equation

$$J = \frac{1}{2} \sum_{(i,j) \in S} e_{ij}^2 = \frac{1}{2} \sum_{(i,j) \in S} \left(r_{ij} - \sum_{s=1}^k u_{is} \cdot v_{js} \right)^2$$

- ightharpoonup r_{ij} fixed this is the training data
- lacktriangle u_{is} is the association between user i and one of the k latent factors
 - History / Romance in our example from last week
- \mathbf{v}_{js} is the association between item j and one of the k latent factors
- What is $\frac{\partial J}{\partial u_{is}}$?



Computing $\frac{\partial J}{\partial u_{is}}$



- only user i matters
- can simplify the function

$$= \frac{\partial}{\partial u_{is}} \frac{1}{2} \sum_{j \in I} \left(r_{ij}^2 - 2r_{ij} \sum_{s=1}^k u_{is} v_{sj} + \left(\sum_{s=1}^k u_{is} v_{sj} \right)^2 \right)$$

$$= \frac{1}{2} \sum_{j \in I} \left(-2r_{ij} v_{sj} + 2v_{sj} \sum_{s=1}^{k} u_{is} v_{sj} \right)$$

$$= \sum_{j \in I} \left(-r_{ij} v_{sj} + v_{sj} \sum_{s=1}^{k} u_{is} v_{sj} \right) = \sum_{j \in I} -v_{sj} \left(r_{ij} - \sum_{s=1}^{k} u_{is} v_{sj} \right)$$

$$\blacksquare = \sum_{j \in I} -v_{sj} e_{ij} - \cdots$$

This is our prediction foirmula

 e_{ij} is the error on r_{ij}

Our big hairy gradient



- A bunch of $\frac{\partial J}{\partial v_{sj}}$ terms
 - derivation is very similar
- Now we know the gradient of our loss function at every point
 - that is: we can calculate at each point, what is the steepest uphill direction
 - downhill = uphill
 - we can calculate the steepest downhill direction

+ Example

■ Let's factorize by gradient descent

+ Thursday

- Factorization at a faster pace
 - Add in regularization
 - Alternating least squares