

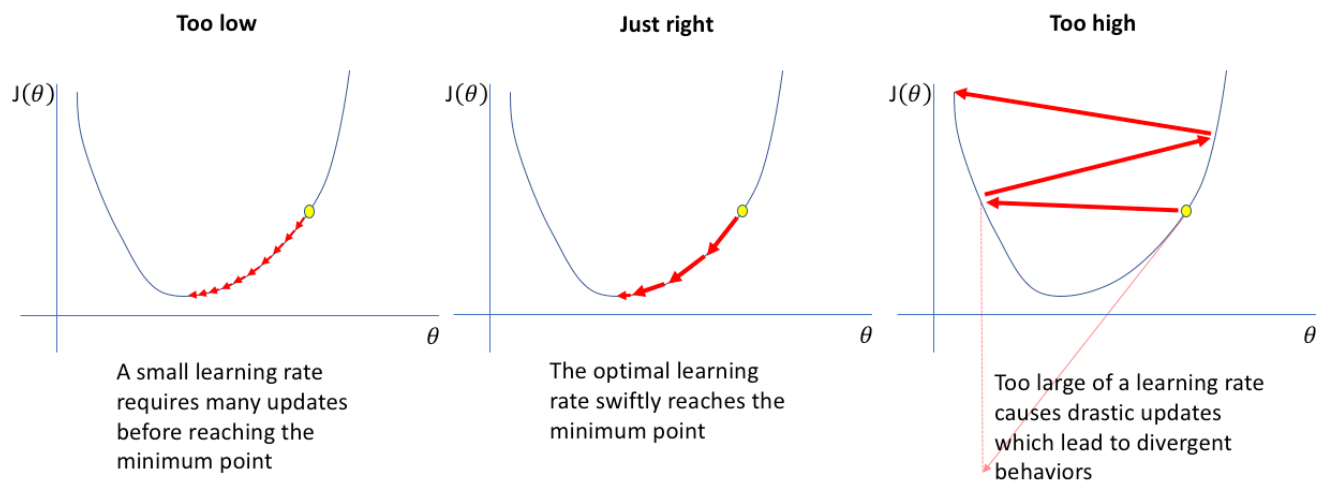
Final Review // Jiyoung Shin

1. What is a language model?

- A language model is a probability distribution over sequences of words. Once any length of words is input, and a next word is output.

2. What is discriminative learning rate and why is it important?

- Discriminative learning rate is to apply different learning rate for different part of the model, which makes training faster.
- If learning rate is too high, it can lead to divergent. If learning rate is too low, it requires many updates.



3. (Tabular)What is normalization for a continuous column?

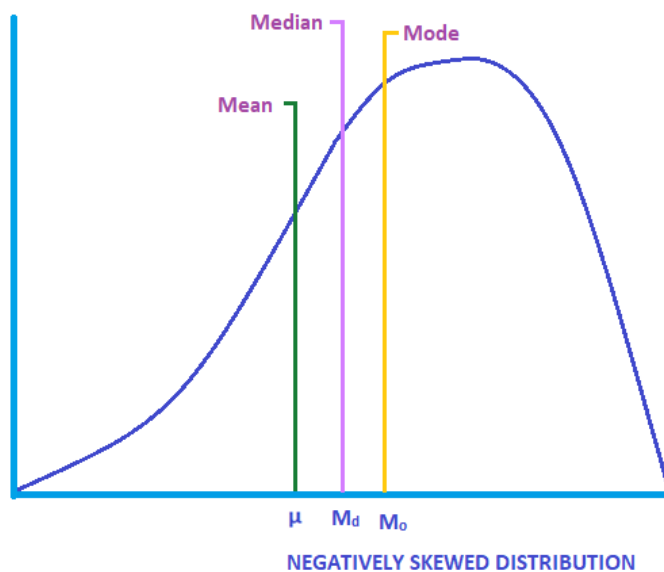
- It is to get the mean and standard deviation of a column.

4. (Tabular)What is a categorical vs a continuous variable?

- **Categorical variables** contain a finite number of categories or distinct groups. Categorical data might not have a logical order. For example, categorical predictors include **gender, material type, and payment method**.
- **Continuous variables** are numeric variables that have an infinite number of values between any two values. A continuous variable can be numeric or date/time. For example, the length of a part or the date and time that a payment is received.

5. (Tabular)How are missing values filled?

- Continuous values can be filled with median or average
- Categorical values can be filled with NULL(0).
- Average vs Median



6. What is collaborative filtering?

- Collaborative filtering is a technique that can filter out items that a user might like on the basis of reactions by similar users.
- Recommendation system

	i_1	i_2	i_3	i_4	i_5
u_1	5		4	1	
u_2		3		3	
u_3		2	4	4	1
u_4	4	4	5		
u_5	2	4		5	2

- It can arise cold start problem which is there is no data.

7. (collaborative filtering). What is the cold start problem?

- The cold start problem is the difficulties posed by the initial lack of meaningful data.

8. What are embeddings? (collaborative filtering)

- An embedding is a mapping of a discrete, categorical variable to a vector of continuous numbers.
- Series of contents which represents object. Longer sequence represent more information.

9-11) $y = mx + b$

9. What is an embedding bias?

- Additional term b .

10. What are activations?

- Output y .

11. What are parameters in machine learning model?

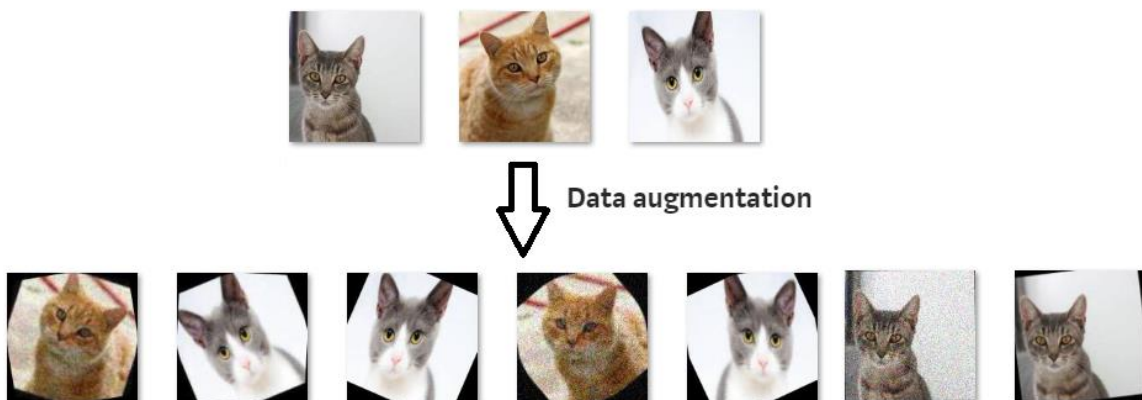
- m and b .

12. What is regularization?

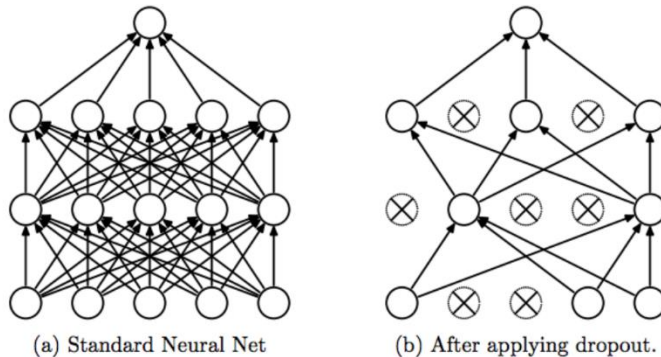
- Adding information in order to solve an ill-posed problem or to **prevent overfitting**.

13. What is data augmentation?

- Data augmentation is to increase amount of data based on only one data.



14. What is dropout?



- Dropout is to delete random nodes, which can provide a very efficient way to train and prevent overfitting.

15. Do you use dropout outside of the training phase?

- No. It's already applied in fastAI

16. What is weight decay?

- Weights are parameters. decay means breakdown. if m and b is big, pull it less than 1.
- Prevent the values blowing up

17. What is an optimizer? What does it optimize?

- An optimizer updates the parameters.

18. What is a type of optimizer?

- ex) Gradient descent, Adam, Adagrad, RMSprop

19. What is momentum?

- You should approach the solution a little bit quicker by momentum
- Momentum in ML is to increase the speed of training.

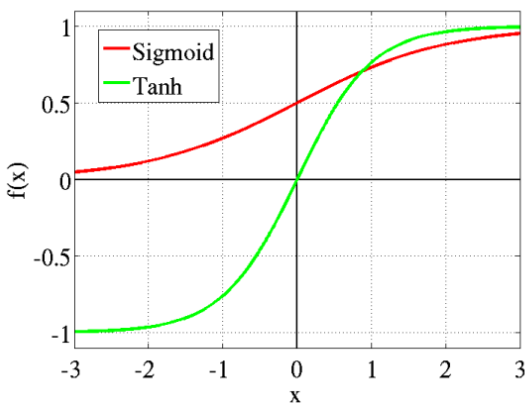
20. What are cyclic learning rates?

- practically eliminates the need to experimentally find the best values and schedule for the global learning rates. Instead of monotonically decreasing the learning rate, this method lets the learning rate cyclically vary between reasonable boundary values.

21. What is softmax?

22. What is a sigmoid function?

- A kind of activation functions.



23. What is categorical cross-entropy loss?

- Use softmax
- https://gombru.github.io/2018/05/23/cross_entropy_loss/

24. What is binary cross-entropy?

25. What is batch norm?

- more easy to train
- expected $y - y_2 = \text{loss function}$.