

## Machine Learning Crash Course – Week 4 Quiz Q&A

### Q1- Which of the following statements is false?

- Selection bias occurs if a data set's examples are chosen in a way that is not reflective of their real-world distribution.
- Sampling bias occurs if proper randomization is not used during data collection.
- Word2Vec exploits contextual information
- Categorical data refers to input features that represent infinite set of choices. ✓

A1- Kategorik verilerde sonsuz küme olmamalıdır.

### Q2- Which of the following statements about online learning are true? Check all that apply.

- I- Online learning algorithms are most appropriate when we have a fixed training set of size (m) that we want to train on.
- II- One of the advantages of online learning is that if the function we're modelling changes over time (such as if we are modelling probability of users clicking on different URLs, and user tastes/preferences are changing over time), the online learning algorithm will automatically adopt to these changes.
- III- When using online learning, you must save every new training example you get, as you will need to reuse past examples to re-train the model even after you get new training examples in the future.
- IV- Online learning algorithms are usually best suited to problems where we have a continuous/non-stop stream of data that we want to learn from.

- I, II
- I, III
- III, IV
- II, IV ✓

**A2-**

- I- Online öğrenme için sabit örnekli bir eğitim mecburiyeti yoktur. Değişken örnek sayılı datasetleri kullanılabilir.
- II- Doğru bir ifadedir.
- III- Eğitim örneklerinin kaydedilip sağlanmasına gerek yoktur, sistem yeni gelen örneklerle kendisini ilgili yöne doğru yönlendirecektir. Bugün geçerli olan bir durum yarın için geçerli olmayabilir.
- IV- Doğru bir ifadedir.

**Q3- Which of the following statement is false about embeddings?**

- A sparse vector which has values 0 and 1 is an embedding. ✓
- The main purpose of an embedding is to find nearest neighbors in the embedding space.
- Typically, an embedding is a translation of a high-dimensional vector into a low-dimensional space.
- An embedding is a mapping of a categorical variable to a vector of continuous numbers.

**A3-** Embeddingler one-hot encoding gibi sparse vectorlerden oluşmazlar, 0 ile 1 arasında continuous değerler alan dense vectorler ile temsil edilirler.

**Q4- Which of the following is more likely to happen to Steve's profession?**

An individual has been described by a neighbor as follows: "Steve is very shy and withdrawn, invariably helpful but little interest in people or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." (Daniel Kahneman, *Thinking Fast and Slow*)

- Librarian
- Farmer ✓
- Politician
- CEO

**A5-** Steve'in özellikleri bir kütüphaneciyi anımsatması bir yanlışlığa sebep oluyor. Ancak Steve'in bir çiftçi olması daha olasıdır. Çünkü toplumdaki çiftçi oranı diğer mesleklerin oranından oldukça fazla olduğundan bu özelliklere uyan çiftçilerin olması daha olasıdır.

**Q5- Which of the following statements are true for dynamic training?**

- I- Very little or no monitoring of input data need to be done at inference time
- II- Very little or no monitoring of training need to be done
- III- Our model stays updated as new data arrives
- IV- Being stale is not a problem of dynamic training

- All of them
- I, III
- III, IV ✓
- I, II
- None of them

**A5-** Dinamik eğitimde model güncel kalır ve eskime/bayatlama (staling) dinamik eğitim için bir problem değildir, dolayısıyla III ve IV seçenekleri doğrudur.

**Q6- Which one is a metric for when we don't focus on recall or precision, and try to find the balance between two?**

- Accuracy
- F1-Score ✓
- p-value
- Loss

**A6-** Precision'la recall'u dengeleyecek bir noktayı bulmak için F1-Score hesaplarız.

**Q7- Which of the following statements are true for static training?**

- I- You can verify the model before applying it in production
- II- The model does not stay up to date as new data arrives
- III- Static training requires less monitoring of training than dynamic training
- IV- We can use batch training and testing, iterate it until model is good

- All of them ✓
- I, III
- III, IV
- I, II
- None of them

**A7-** Tüm ifadeler static training için doğrudur.

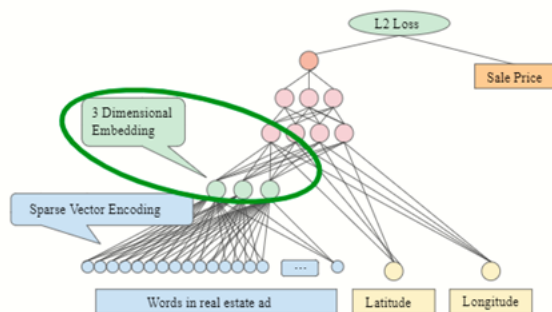
**Q8-** Which one of the following is a classification problem that we should focus on precision value?

- Classifying people as eligible for a credit (positive) or not (negative)
- E-mail spam detection ✓
- Cancer diagnosis clinic which classifies patient as positive and negative
- Classifying a dog as cute or not where the cost of false classifications are low

**A8-** Spam olmayan bir e-maili spam olarak işaretlemenin (false positive) maliyeti büyük olacağından spam detection'da precision'a bakılır.

**Q9-** In embeddings, suppose you have 3 units in the embedding layer as shown below. In this case, how would you represent those embeddings and which type of propagation would you use to train the proposed model?

Regression problem to predict home sales prices:



- 1 dimensional representation and feed forward propagation
- 2 dimensional representation and back propagation
- 2 dimensional representation and feed forward, back propagation
- 3 dimensional representation and feed forward ,back propagation ✓

**A9-** Embedding (gömme) katmanı 3 tane nörondan oluştuğu için, 3 boyutlu bir şekilde gösterebiliriz. Örnek olarak film tavsiyesinde 2 boyutlu eksen üzerinden gösterilmiştir ve orada 2 tane nöron mevcuttur. Bu tür modeller oluşturmak için hem ileri yönlü hem de geri yönlü yayılım kullanılmaktadır.

**Q10- Which of the following statements are true for dynamic inference?**

- I- We can do post verification of predictions before they are used
  - II- We can provide predictions for all possible items
  - III- Cost of inference can be more problematic in static inference with respect to dynamic inference
  - IV- Prediction latency can be more problematic in dynamic inference with respect to static inference
- 
- All of them
  - I, III
  - III, IV
  - II, IV ✓
  - None of them

**A10-** Gelen tüm girdiler için tahmin çıkarımını yapabiliriz ve tahmin gecikmesi dinamik çıkarımda statik çıkarıma çok daha büyük bir problemdir, dolayısıyla II ve IV doğru olan ifadelerdir.