

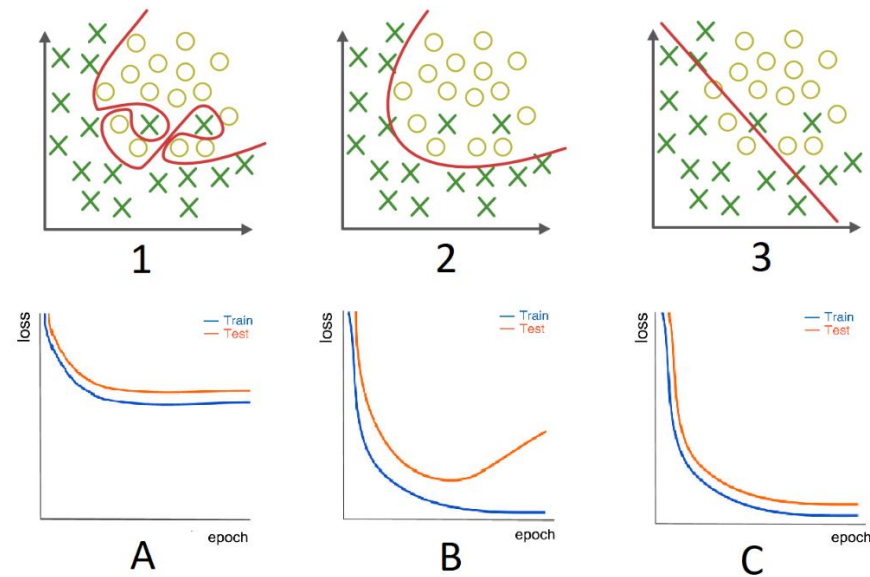
Machine Learning Crash Course – Week 1 Quiz Q&A

Q1- Which of the below is the reason why we are using validation sets?

- The regressor might overfit to test set if we don't use validation sets. ✓
- Validation sets increase model's performance on the fit.
- Validation sets help us check if the model overfits or underfits.

A1- Bu soruya veriye validation set eklemesek ne olur'la baksak çok daha iyi olacak. Siz bir model geliştiriyorsunuz, training ve test seti ayırdınız, 100 örnekten 80'i training 20'si test. Ev fiyatlarını tahminlemeye çalışıyorsunuz. Bir regresyon modeli train ettiniz, sonra içine test verisinden 4 odalı ve iki banyolu bir ev koydunuz o da size bu evin fiyatının 100 bin lira olması gerektiğini söyledi, ama gerçekte o ev (test verisindeki ev fiyatı kolonu) 120 bin lira, buna göre hatanıza baktınız, parametrelerinizi değiştirip yeniden train ettiniz. Zamanla kendinizi bu test verisinden aldığınız hatalara göre adapte ediyorsunuz, yani test verisine overfit ediyorsunuz. Farkettiyseniz test verisiyle hem parametreleri değiştiriyoruz hem de test ediyoruz, bu yanlış, bu yüzden validation set ekliyoruz, hataya bakıp parametre değiştirme işlemini validation set'te yapıyoruz, ardından yeni çıkan modeli test verisiyle test ediyoruz, böylece modelin gerçekten iyi bir performans sergileyip sergilemediğini görebiliyoruz.

Q2- You can see different models to classify "X" or "O" for a dataset below. Which loss-epoch graph would you expect for these models.



- 1-B / 2-C / 3-A ✓
- 1-C / 2-A / 3-B
- 1-A / 2-B / 3-C
- 1-B / 2-A / 3-C

A2- Soruda "X" ve "O" olmak üzere iki adet sınıfı bulunan bir sınıflandırma (classification) problemi için 1, 2 ve 3 numaralı grafiklerde görülen farklı modeller train edilmiş. Bu modellerden train ve test dataları için nasıl bir loss-epoch grafiği beklendiği soruluyor.

- 1 numaralı modelimize baktığımızda train datası üzerinde hatasız bir şekilde eğitildiğini görüyoruz. Aşağıdaki resimde test datasının siyah renk ile gösterildiğini varsayalım. Bu model test datasında sınıflandırmaları hatalı yapacaktır, çünkü train datasını çok fazla öğrenmiştir yani bu modelde "overfitting" gerçekleşmiştir. Beklediğimiz loss-epoch grafiği "B"deki gibi olmaktadır çünkü train datasında loss oldukça düşük iken test datasında loss artmaktadır.
- 2 numaralı modelin birkaç hata dışında datamızı öğrendiğini ve basit bir eğri ile iyi bir sınıflandırma yapabildiğini görüyoruz. Burada 1 numaralı model gibi bir komplekslikten ve bu kompleksliğe bağlı hatalardan söz edemeyiz dolayısıyla train ve test dataları için alınan loss değerlerinin birbirine yakın olacağını öngörebiliriz. 2 numaralı modelin grafiği "C" olmaktadır.
- 3 numaralı modelde ise sınıflandırma çok basit bir seviyede tutulmuş ve hatalar diğer iki modele göre çok daha fazla görünüyor. Modelin yeterince eğitilmediği bu gibi durumlara ise "underfitting" diyoruz. Bu durumda 3 numaralı modele karşılık gelen grafik "A" olacaktır.

Q3- A dataset with 250 samples is split into training and test set in an 80-20 ratio. The model was trained using Stochastic Gradient Descent using the mini-batch size of 32. After successful 1000 epochs, how many times the model was updated during training?

- 32
- 6000
- 200
- 7000 ✓

A3- $250 * 0.8 = 200$ eğitime girecek veri sayısı. 200 adet örnek 32'lik batch'lere ayrılıyor. Toplam batch sayısı 6 ancak 8 adet örnek dışarıda kalıyor. Bunlar birlikte batch oluşturduklarından toplam batch sayısı 7 olur. Yani bir epoch tamamlanana kadar modelin güncellenme sayısı 7. 1000 epoch sonucunda model 7000 kere güncellenecektir.

Q4- Which of the below wouldn't be a good feature for when predicting house prices?

- Location of the house
- Square footage of the house
- Previous owner's gender ✓
- Number of bedrooms

A4- Önceki ev sahibinin cinsiyeti modeli geliştirmede kullanılacak özelliklerden biri değildir, alakasız bilgidir.

Q5- Which of the following is not a problem caused by not being able to choose the optimal learning rate?

- It takes long time to reach minimum point of the loss curve.
- Taking a long step without noticing the minimum point of the loss curve.
- Taking a step towards the wrong direction for to reach minimum point of the loss curve. ✓

A5- Learning rate, modelimizi eğitirken "Ne kadar öğreneceğim?" sorusuna cevap olarak verilebilecek bir katsayı olarak karşımıza çıkar. Bu yüzden bu katsayının çok yüksek seçilmesi loss functionda en düşük durumu görmeden atlayabilme ihtimali yaratırken çok düşük seçmemiz ise eğitim süresinin çok uzun sürmesine sebep olacaktır. Son şık ise gradyanın yanlış yönlü alınması ile gerçekleşebilecek bir sorun olarak karşımıza çıkmaktadır.

Q6- Which of the following options is a linear regression problem?

- Predicting the comment on the e-commerce site is negative or positive
- Predicting the gender of a person by his/her handwriting style
- Predicting author of a text
- Predict the sales of the number of copies a music album for next year ✓

A6- Sınıflandırma benzer içeriğe sahip verilerin kategorize edilme işlemidir. Sınıflandırmada grup üyeleri tanımlanır, veriler bazı parametrelere göre farklı etiketler altında kategorize edilir ve ardından veriler için etiketler tahmin edilir. Regresyon ise sınıflandırmadan farklı olarak geçmiş ve mevcut verileri kullanarak bir etiket yerine sayısal veri tahmini için kullanılır. Sorumuzun şıklarını bu bilgiler ışığında ele alacak olursak bir metnin yazarının tahmini, bir e-ticaret sitesindeki yorumların negatif mi yoksa pozitif mi olduğunun tahmin edilmesi ve el yazısı stiline bakarak cinsiyet tahmini bir sınıflandırma problemi iken gelecek yıl için müzik albümü kopyalarının satışının tahmini ise bir regresyon problemidir.

Q7- Which one is true related to the "correlation matrix"?

- The lower the absolute value of a correlation value, the greater its predictive power.
- It indicates how each attribute's feature values relate to the other attributes' raw values.
- It indicates how each attribute's raw values relate to the other attributes' raw values. ✓
- "0.0" means no correlation; the two columns are linearly correlated.

A7- Bu soru, kursun ilgili Tensorflow bölümünde ayrıntılı olarak açıklanmaktadır. Öğrendiklerimizi hatırlayacak olursak; matrisin mutlak değeri ne kadar yüksek olursa, hedeflenen değeri de o derece tahmin edebilir olur. Bu CMatrix tanımından gelir. Onun temel özelliği ham değerler arasındaki benzerlikleri tespit etmektir. Yani doğru cevap budur. 0.0, korelasyon olmadığı anlamına gelir fakat onlar arasında lineer olarak da korelasyon yoktur.

Q8- Which of the following is prevent overfitting?

1. Cross-validation
2. Training with more data
3. Removing features
4. Early stopping
5. Regularization
6. Ensembling

- 1,2,4,5
- 1,2,5,6
- 2,3,4,5
- All of them. ✓

A8- Modelin ezberlemiş olduğu durumlara genel olarak overfitting denir. Bunu engellemenin yöntemlerinden bir kaç aşağıdaki gibidir.

- Cross validation: Elimizdeki data az ise, her seferinde elimizdeki datanın bir kısmını test datası olarak belirleyip, böylece her datanın en az bir kez test datası olarak kullanılması sonucu yapılan çalışmadır.
- Training with more data: Datanın yetersiz olması ve eğitim datasının test datasından uzak olması durumunda geçerlidir.
- Removing Features: Bir çok özelliğin olması durumunda modelin tüm özellikleri ezberlemesi sonucu yapılan çalışmadır.
- Early stopping: Modelinizin eğitimini uygun yerde durdurulmaması sonucu yaşanan durumlarda yapılması gereken çalışmadır.

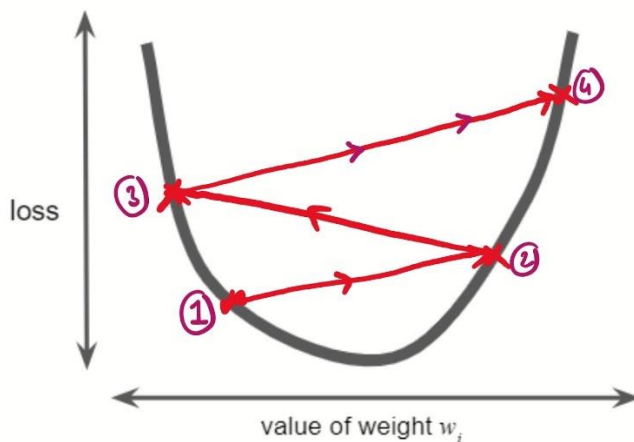
- Regularization: Öğrenilmiş bir modeli daha geniş veri kümelerini kapsamaları için, yani genelleştirme için maliyet fonksiyonuna pozitif etki olarak eklenen cezalandırma işlemine denir.
- Ensembling: Birden farklı tahmin sisteminden faydalananak verilen tahminlerin ortalaması olarak karar veren bir çalışmadır.

Q9- Which one is the unnecessary assumption when applying test methodology to test the fit?

- The examples are independently and identically drawn from the set.
- The distribution of the set doesn't change, no new examples are added.
- The examples should always be drawn from the same set.
- Randomization should be done when sampling the examples.
- The model should be trained on the test data. ✓

A9- Test verisi training verisinden birbirinden bağımsız ve rastgele seçilmiş olmalıdır ki yanlılık olmasın, mesela kayak malzemesi satılan bir şirketin satış verisi üzerinden tahmin yaparken sadece kış verisini almak yanıltıcı olacaktır. Aynı şekilde test setinin training setini yansıtmaması adına sürekli aynı training setinden örnekler alınmalıdır, dışarıdan ekleme yapılmamalıdır. Model test verisi üzerinden eğitilmemelidir.

Q10- The loss value is increasing in four iterations in the figure below while training a model. Which parameter value should be chosen for a lower value for decreasing the value of loss?



- Learning rate ✓
- Epoch
- Batch size

A10- Şekilde görüldüğü gibi iterasyon sırasında loss(hata) değeri azalması beklenirken artmış, ayrıca loss(hata) değeri o kadar yüksek artmış ki dip noktayı ıskalamış(overshooting). Hata fonksiyonunda her bir iterasyon sırasında gradyant operatörünün genliği ve yönüne adım atılıyordu. Öğrenme katsayısı (learning rate) atılan bu adımın büyüklüğü yani gradyant operatörünün genliği ile doğru orantılıdır. Şekilde gösterilen hata değerinin bu şekilde artması, öğrenme katsayısının gereğinden çok büyük olmasıdır. Diğer bir ifadeyle Goldilocks öğrenme katsayısına göre gradyant operatörü güncellenmemiştir.