

10:37:21 From 20175499 이성준 : 텐서에서 dot 연산의 의미가 ?

> Please see this page (<https://numpy.org/devdocs/reference/generated/numpy.dot.html>)

10:38:08 From 노운영 4076 : 프로젝트는 개인폴젝예정인가여 그룹폴젝예정인가요?

> Both projects are individual projects.

10:39:00 From KAIST SoyeonJung : ML basic에 관한 추천해주실 책이나 material이 있으신가요?

> I recommend reading Pattern Recognition and Machine Learning by Bishop for building fundamental understanding of machine learning. As for basic machine learning, I'm sure there are tons of material on Udemy, Udacity or Stanford online courses.

10:40:13 From 20203640 최별이 : final quiz 예시 답으로 공유해주셨던 코드에서 M, N 값 설정할때 10을 빼고 나누는 이유가 궁금합니다

> I was just trying to keep the M and N to small values for nicer output.

10:41:24 From 20203344 신현호 : 어제 저녁에 수강 정정해서 들어왔는데 Quiz가 있었나요..?

> There is no quiz in this course. Even if there is, it is not counted towards the final grade.

10:42:57 From 심상민 (20203347) : final quiz에서 row-wise mean이 있었는데 axis=0으로 해야하는게 맞는건가요?

> Please disregard the word "row-wise mean". I was just trying to make the quiz a bit more difficult, but in retrospect, I should not have chosen that term, because it can be confusing.

11:03:41 From 20204871 YH Tan : meta learning is supervised learning?

> Meta learning is "learning to learn", so you can consist of the meta rounds with supervised learning for unsupervised learning.

11:04:50 From Rushda Basir : non-parametric techniques such as k-NN unsupervised learning technique?

> You still need labeled samples to classify a given input via k-NN, so it is supervised method. But since you are not "learning" any parameters, I'm not sure if you call it supervised learning.

11:06:40 From 20204871 YH Tan : how about transfer learning?

> Transfer learning is similar to meta learning. If you train your VAE on ImageNet, then transfer to SVHN, then you are transferring an unsupervised learning model.

11:07:53 From Rushda Basir : how about NLP for speech generation?

> If you are generating something (e.g. images, speech, text), you can think of it as both supervised learning and unsupervised learning.

11:16:50 From Kyeonghyeon Park : - 까지가 cross entropy loss인가요?

11:17:58 From young jun Choi (20204581) : 네

11:23:33 From 20203162 김진호 : unsupervised learning의 경우

11:23:40 From 20203162 김진호 : y가 없는데 loss function은 어떤식으로

> In unsupervised learning, there is some loss function you need to minimize. If you are training VAE, you are using reconstruction loss. If you are doing clustering, you can use how intra-cluster members are in close distance to one another and inter-cluster members are far away from one another.

11:25:30 From 변지영 : 선택하는 loss function에 따라서 모델의 퍼포먼스가 많이 달라지나요? 아니면 rule of thumb이 있는건가요?

> There is usually rule of thumb. For classification, you use cross entropy, or negative log likelihood. For regression, you use mean squared error.

11:28:14 From Kyeonghyeon Park : 말씀하신 auto encoder 예시의 경우 자기 자신이 label인 supervised learning이라 생각하면 오류가 있는건가요?

> I think it is reasonable to view autoencoders like that.

11:30:21 From 51560 김성민 : clustering이 얼마나 잘되었는지 서로 다른 model간의 비교를 할 수 있는 대표적인 방법이 있나요?

> If you already know the true membership of each sample, then you can use that information to evaluate the quality of your clusters. If not, you can use heuristic methods such as the distance between intra-cluster members and inter-cluster members.

11:34:06 From 박성진 20204341 : autoencoder로부터 복원된 픽셀마다 loss가 다를것같은데, loss가 큰부분을 anomaly point라 짚는다고 한다면 어디까지 anomaly로 볼것인가에대한 threshold를 어떻게 지정할수있는건가요?

> I'm not sure if you can use autoencoders to detect pixel-level anomaly. I think you can use autoencoders to detect sample-level anomaly though. (i.e. If some sample has poor reconstruction, then it's probably because the autoencoder has not seen such sample before)

11:34:23 From yoomi baek : 만약 unsupervised learning에서 label(class)을 만들고, 클러스터링을 할 수 있나요?

> If you already know the true label (membership) of each sample, then you can use them to evaluate the quality of your cluster.

11:41:48 From 20203375 오동원 : f가 해석함수라면 극점들중에 항상 global minimum이 있는것아닌가요? 네트워크에 max pooling같은 비해석 activation이 없으면 극점으로도 global minimum을 구할 수 있는 방법이 있나요?

> It is true that typical analytical functions always have a global minimum. However, finding the global minimum is a different story. If you are dealing with super high-dimensional functions (i.e. Transformers), then there can be millions of local optima, and it is impossible to find which one is the global minimum, unless you have infinite amount of time and computing resources.

11:42:12 From jaegyunkim : partial derivative로 최적의 theta 찾을때,
해당 partial derivative는 그 point에서 기울기라 + partial derivative가 optimal pt로 갈지
- partial derivative가 optimal pt로 갈지 모르는것 아닌가요? $\theta := \text{partial derivative} * \text{learning_rate}$ 로 하는게 항상 optimal pt로 가는게 궁금합니다. + partial derivative가 오히려
반대의 방향일수도 있지 않을까 해서요

> You are right that, in order to find the global minimum, you have to risk going against the gradient. However, in deep learning, you are usually satisfied with finding a local minimum, so you rarely go against the gradient.

11:49:40 From Rushda Basir : how do we determine the batch size?

> That depends on a lot of factors including your model size, the task you are solving, and how powerful your hardware is. For example GPT3 used 3 million samples in a single minibatch.

11:50:36 From 이윤지 20203489 : Could you explain about I.I.D once again?

> I.I.D = Independently, identically distributed: You are assuming that a set of samples come from the same distribution, and they are sampled independently.

11:51:18 From Mario Choi : Should we interpret "SGD -> GD" to mean that SGD will converge on a local minimum, or that θ_{SGD} yields similar results to θ_{GD} ?

> I should say that, if your function has one local minimum (i.e. global minimum) then SGD will be able find the same solution as GD, if we run SGD long enough.

11:55:52 From Rushda Basir : what is the use of validation set?

11:57:47 From Seung Jun Lee : I know validation set is used to decide the optimal hyperparameter like the value k in k-NN

> You also use the validation set to know when to stop training your model. If your validation error starts to increase while your training error keeps decreasing, then it is a good idea to stop your SGD process.

11:57:56 From 유명성 : model capacity가 어떤걸 의미하는 거죠?

> Model capacity is the same as how powerful your model is. A 100-layer neural network is more powerful (i.e. has a higher capacity) than a 10-layer neural network.

12:14:45 From 석원 이 : how can we find an optimal lambda value?

> For regularization, you have to try out multiple lambda values by evaluating on the validation set. (Never use the test set for choosing the hyperparameters!)

12:21:05 From 51560 김성민 : SVM은 binary classification에서만 사용가능한가요?

12:25:09 From Mario Choi : binary classification을 multiclass classification으로 확장할 수 있습니다. 가장 간단한 extension은 여러번 학습해서 class 1 vs non-class1, class2 vs non-class2 등. 그외 방법들은 wikipedia에 (support vector machine 페이지에 multiclass SVM 섹션에) 나오네요.

12:25:53 From 박민영(20203758) : N fold cross validation is only for selecting hyperparameters, and not for determining weights?

> N-fold cross validation is not for selecting hyperparameters. It's for reliably evaluating the generalization performance of your model.