# Q1

* H(x) = sigmoid ( theta dot multiply x )

Range is [0,1]

H(x) is interpreted as the probability of the predicted sample to be good condition

* Cross-entropy is positive
* We can find that for good working condition, ( 1 – y ) = 0. So if h(x) is high value, it will lead to little loss contribution. On the contrary, if h(x) is low value, it will lead to large loss contribution.
* We can fiind that for poor working condition, y = 0. So if h(x) is high value, it will lead to large loss contribution. On the contrary, , if h(x) is low value, it will lead to little loss contribution.

# Q2

* import numpy as np

output = list((-3, 2, 4, 1, 7, -5, -1, -3, 3, 5))

output = np.array(output)

exp\_output = np.exp(output)

denominator = np.sum(exp\_output)

softmax = exp\_output / denominator

softmax

result:

array([3.74251197e-05, 5.55438024e-03, 4.10416272e-02, 2.04334230e-03,

8.24343118e-01, 5.06493917e-06, 2.76536309e-04, 3.74251197e-05,

1.50983709e-02, 1.11562709e-01])

* The each result in output of Softmax function represents the probability of corresponding class to be true.
* In practice, when we need to do multi-class classification, we need to get all the probability for each class. So we can apply Softmax function to convert the predicted value from each class to be probabilistic form to select the true class.

SoftMax Function does two things:

Transform the vector into probabilities that are between 0 and 1 and add up to 1 Transform the largest value into a value close to 1 and the lowest into one close to 0

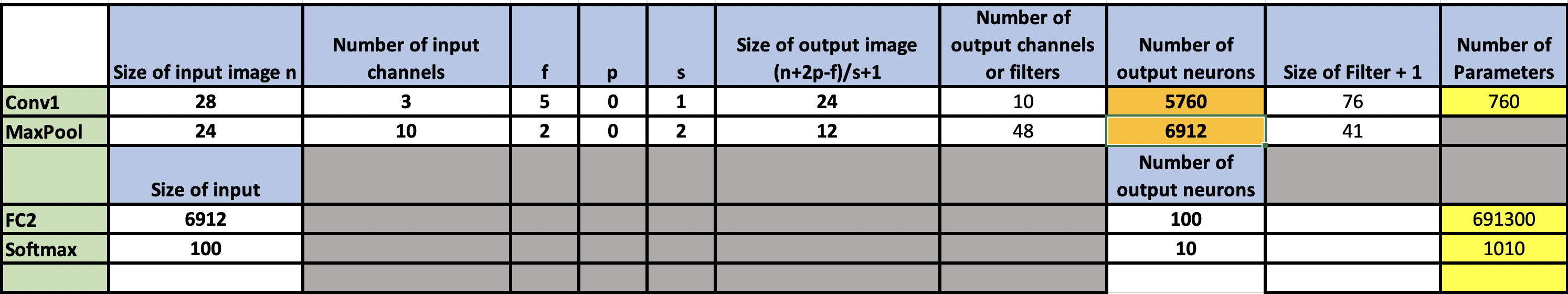
# Q3

* Do not need many training data for new task

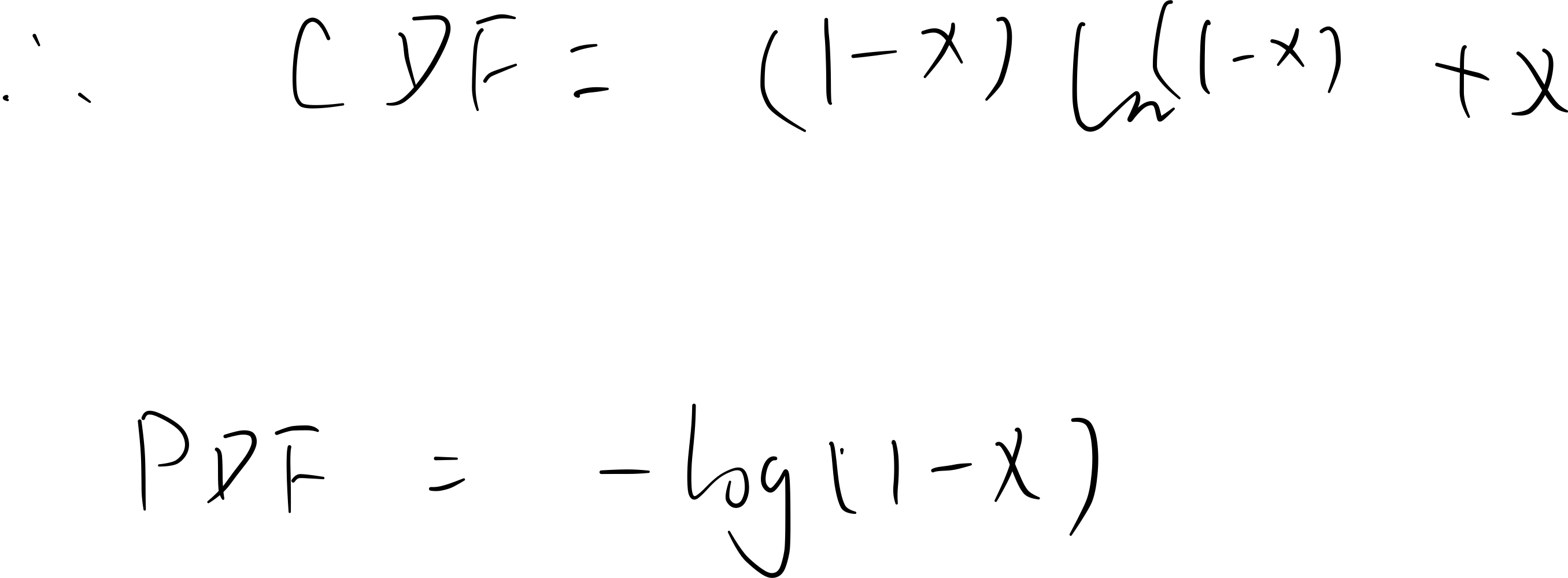
Save much time to train all classifier

* ImageNet is better because the car and the plane in this task are subclass of ImageNet rather than that in MNIST. So the features in classifier trained from ImageNet is appropriate for this task. But features in classifier trained from MNIST is just gray-image and features are for digits.
* Learn rate, momentum and weight decay in Optimizer
* Grid Search method because we only need to train for the last layer for the pretrained classifier. So the time consuming in each epoch is small and it is appropriate to apply grid search method.

# Q4



# Q5



# Q6