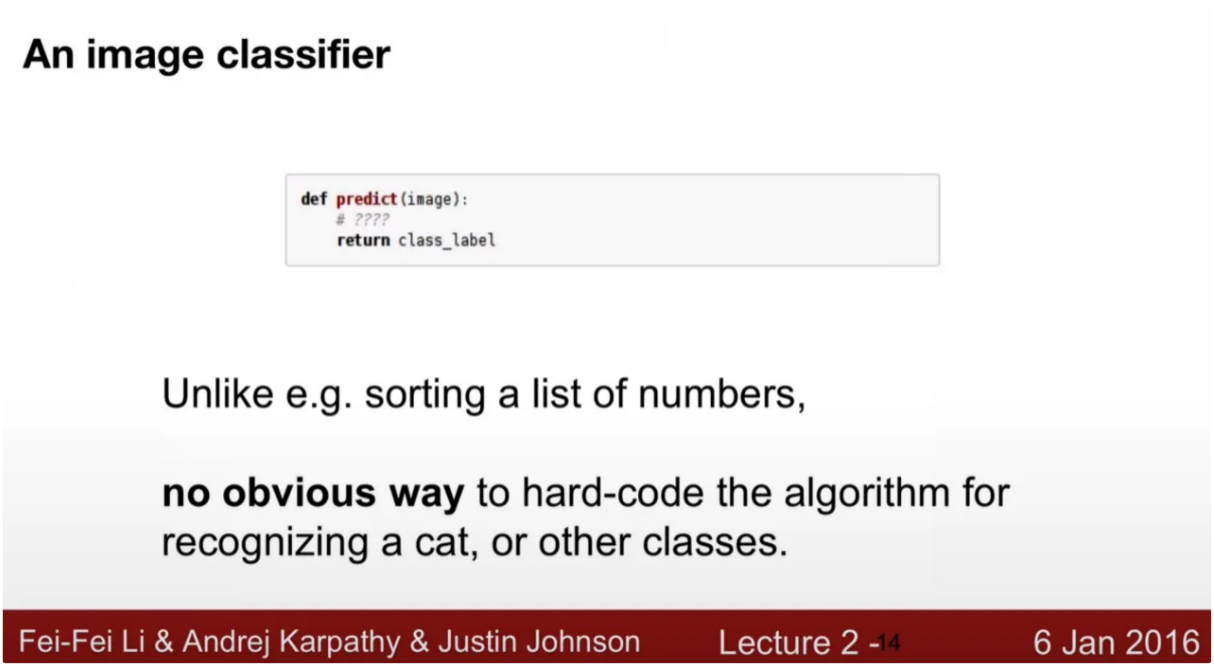
exam 개인학습

Lec 2

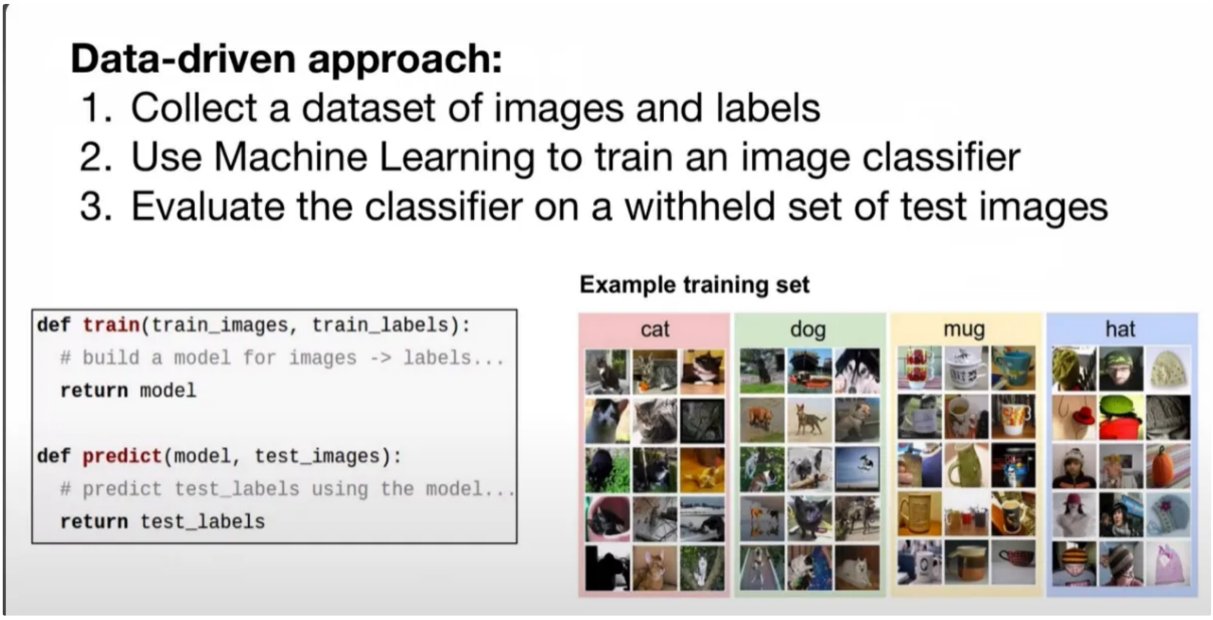
Type something...

01) Image Classification

01_Image Classification

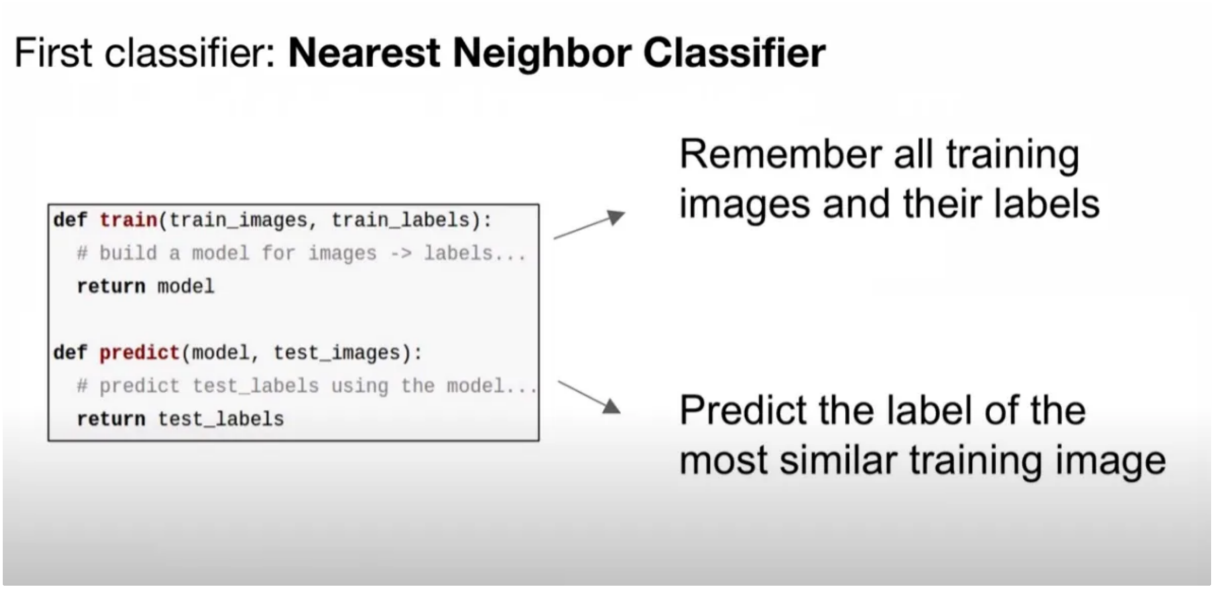


- Image를 분류하는 hard-code algorithm은 없다.



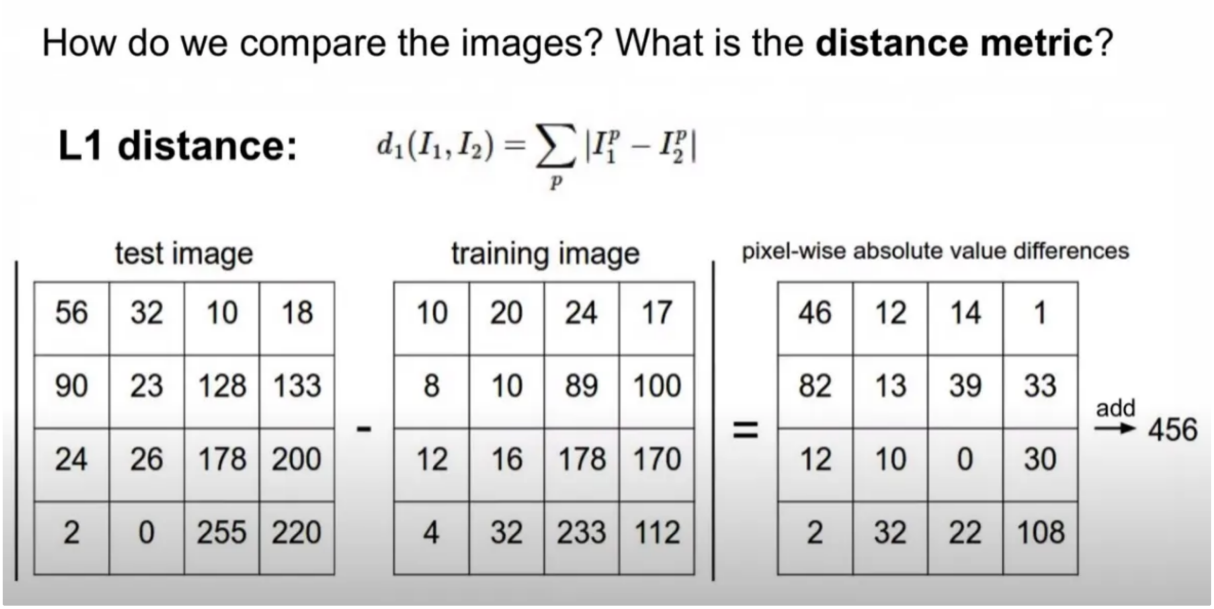
- 그래서 Data를 학습해서 값을 예측하는 접근을 시도하였다.

02_Nearest Neightbor Classifier(KNN)



- 모든 training image를 비교해서 가장 가까운 것에 label를 사용한다.
- 현재는 아무도 사용하지 않는 알고리즘이다.

03_DIstance Metric



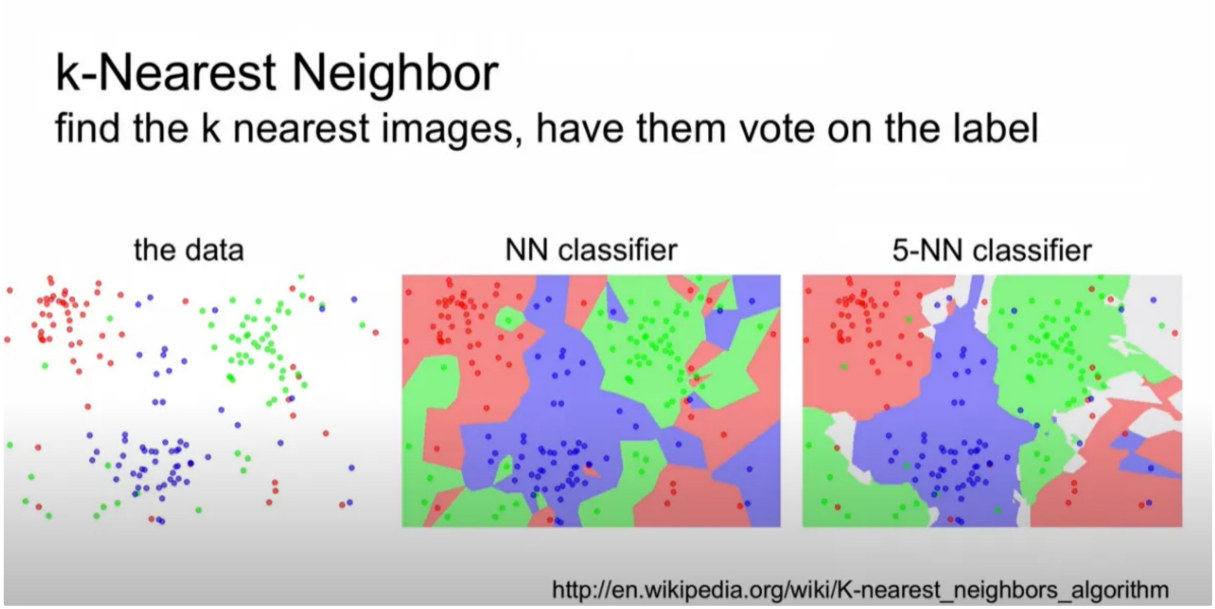
- test image - training image의 절대값을 씌워준 형태를 사용한다
- L1 distance를 통해 구한 거리 중에서 가장 작은 값이 그 test_image의 label이 된다.

Generated by Notion to PDF

```python
import numpy as np

class NearestNeighbor:
  def __init__(self):
    pass

  def train(self, X, y):
    """ X is N x D where each row is an example. Y is 1-dimension of size N """
    # the nearest neighbor classifier simply remembers all the training data
    self.Xtr = X
    self.ytr = y

  def predict(self, X):
    """ X is N x D where each row is an example we wish to predict label for """
    num_test = X.shape[0]
    # lets make sure that the output type matches the input type
    Ypred = np.zeros(num_test, dtype = self.ytr.dtype)

    # loop over all test rows
    for i in xrange(num_test):
      # find the nearest training image to the i'th test image
      # using the L1 distance (sum of absolute value differences)
      distances = np.sum(np.abs(self.Xtr - X[i,:]), axis = 1)
      min_index = np.argmin(distances) # get the index with smallest distance
      Ypred[i] = self.ytr[min_index] # predict the label of the nearest example

    return Ypred
```

Nearest Neighbor classifier

- train 과정이 단순히 데이터를 기억하기만 하기 때문에 시간이 매우 짧다. O(1)
- predict에서는 X를 모두 다 돌아보기 때문에 O(N)
- We want to **Train slow, Predct Fast**



- K가 커질수록 smooth하다.

Generated by Notion to PDF

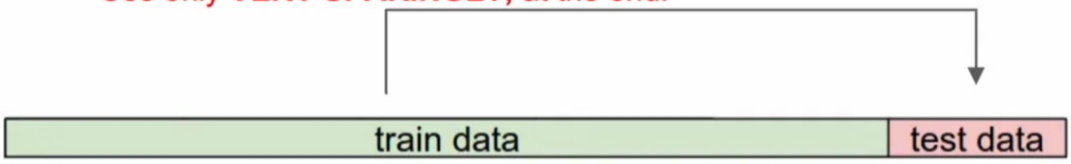
Hyperparameters

What is the best value of **k** to use?
What is the best **distance** to use?

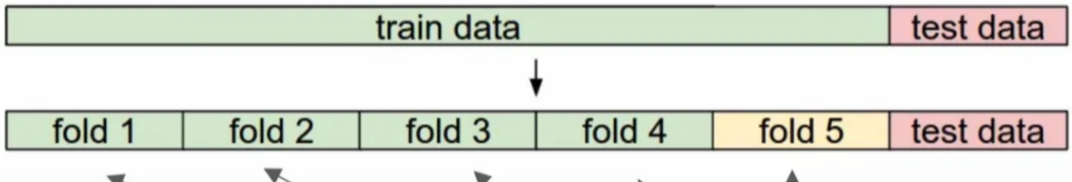
These are **hyperparameters**: choices about the algorithm that we set rather than learn

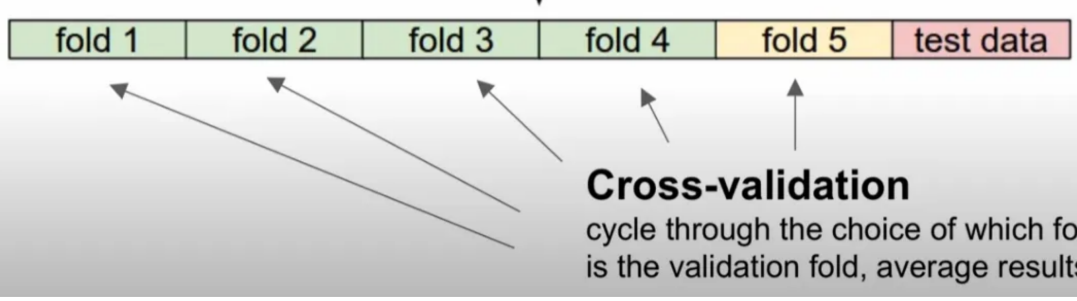
- 가장 최적의 k값, 가장 최적의 distance를 구하는 것들, 그런 것들을 **hyperparameters** 라고 한다.

Trying out what hyperparameters work best on test set:
Very bad idea. The test set is a proxy for the generalization performance!
Use only **VERY SPARINGLY,** at the end.



- test data는 test 때 딱 한번만 사용하는 것이다. 절대로 train 때 사용해서는 안된다.





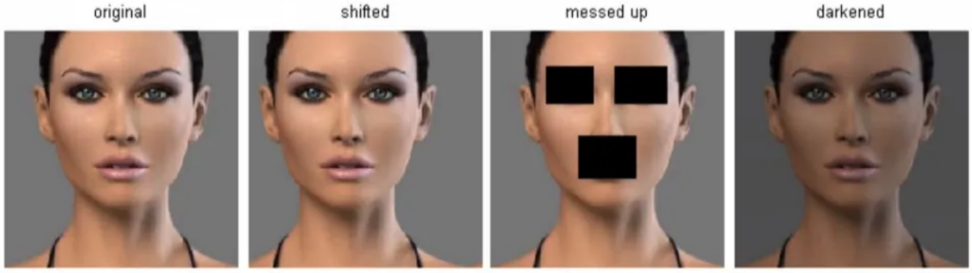
Cross-validation
cycle through the choice of which fold is the validation fold, average results.

- validation set를 포함한 다음, 각각 하나씩 제외하고 남은 하나로 검토하는 **Cross-validation**을 사용하면 매우 좋다.

Generated by Notion to PDF

k-Nearest Neighbor on images **never used.**

- terrible performance at test time
- distance metrics on level of whole images can be very unintuitive



(all 3 images have same L2 distance to the one on the left)

- 절대로 KNN을 image에 사용해서는 안된다.
- test time이 매우 느리고, 픽셀에 대한 거리가 별로 유용하지 않다.

K-Nearest Neighbors: Summary

In **Image classification** we start with a **training set** of images and labels, and must predict labels on the **test set**

The **K-Nearest Neighbors** classifier predicts labels based on nearest training examples

Distance metric and K are **hyperparameters**

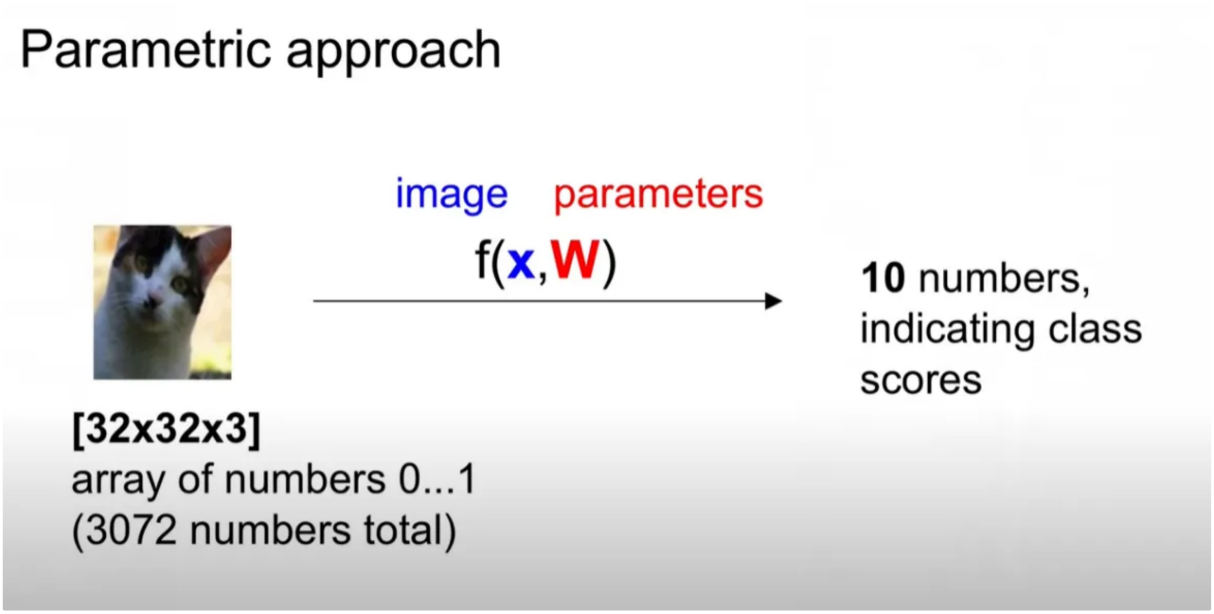
Choose hyperparameters using the **validation set**; only run on the test set once at the very end!

<요약>

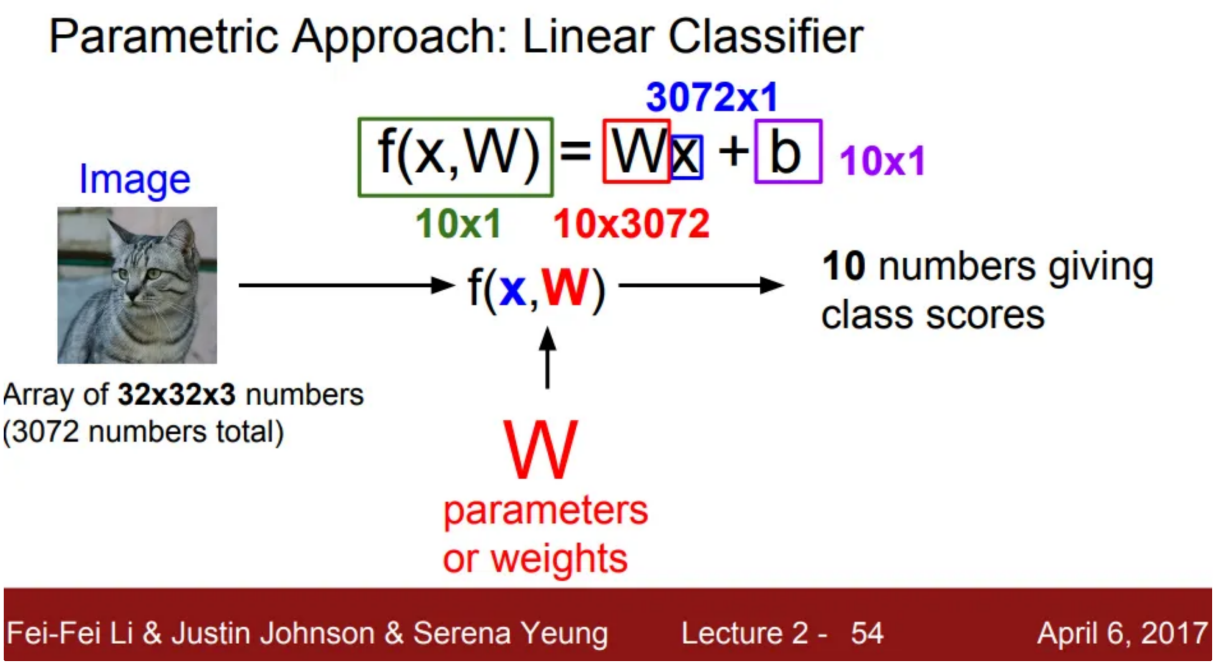
- Image classification에서는 training set과 test set으로 나눈다.
- KNN은 거리를 이용한 predict 방식이다
- Distance metric 과 K값은 모두 hyperparameter이다
- validation set을 사용해서 hyperparameter를 선택한다.

04_Linear Classification

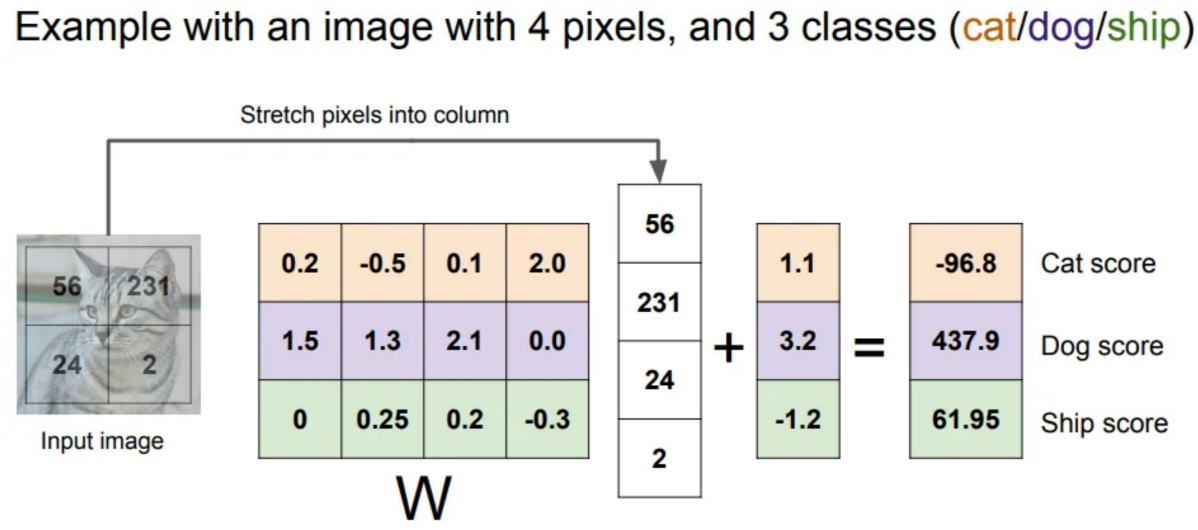
Generated by Notion to PDF



- f(x, W): image + parameters



- f(x, W) = Wx + b
- 10 x 1 = 10 x 3072 * 3072 x 1 + 10 x 1



Generated by Notion to PDF

- 예시는 다음과 같다.

Generated by Notion to PDF