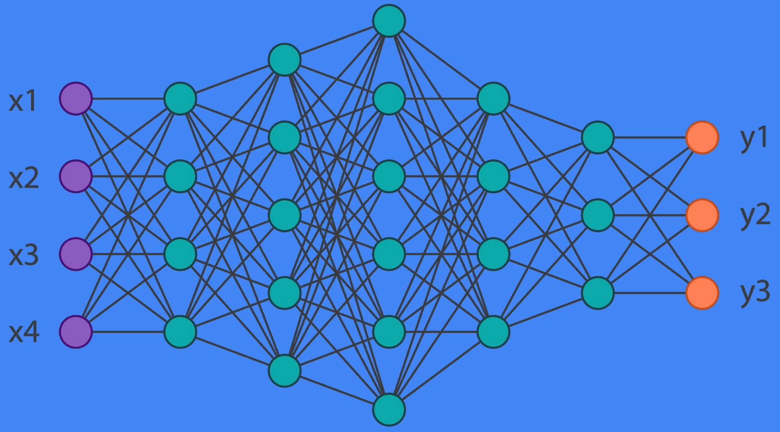



# Neural Networks

Linear Classification and Optimization

# Objective

We want to classify cats and as well as other animals.

We have N training examples, K classes (Number of animals), D pixels (dimension of our picture)

N : 10,000 K : 10 D : m x 64 x 64 x 3

Our objective is map N animals to its specific class

# How we can do it?

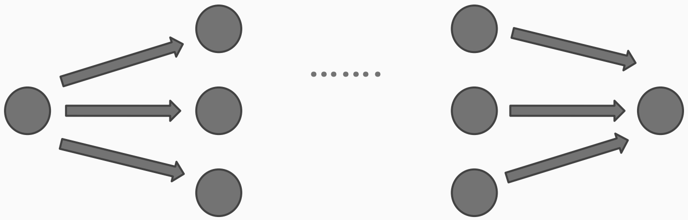
Simplest linear function => $f(x) = Wx+b$

Adding more layers to the networks makes the network recognize patterns and different features

It makes feature engineering for us

# How we can do it?

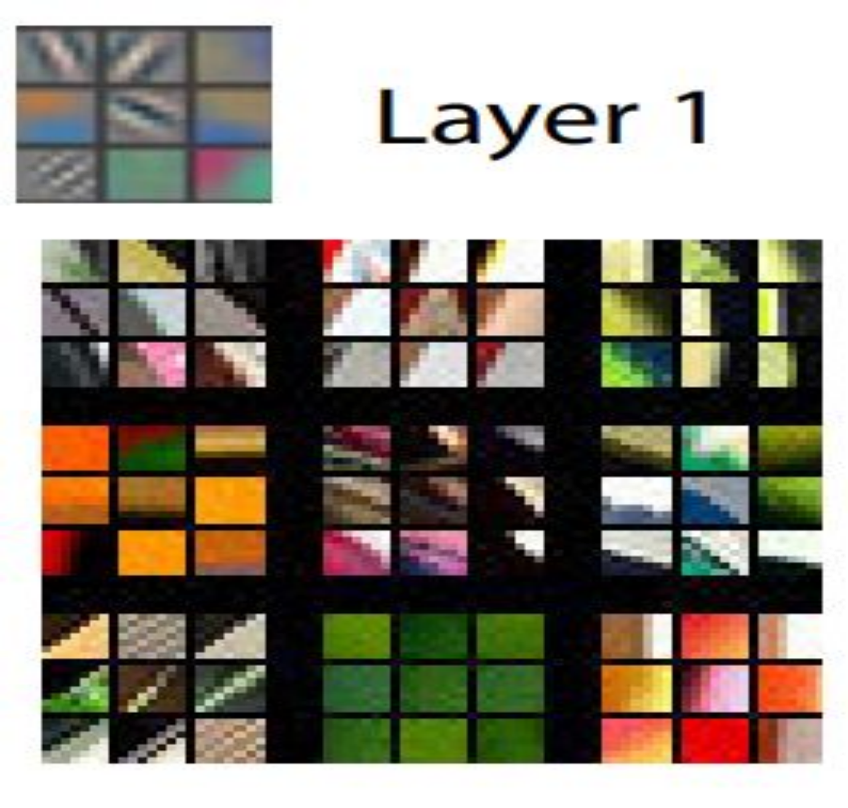
$$f(x) = W_n * (W_{n-1} * (...(W_1 * x + b_1)...) + b_{n-1}) + b_n$$



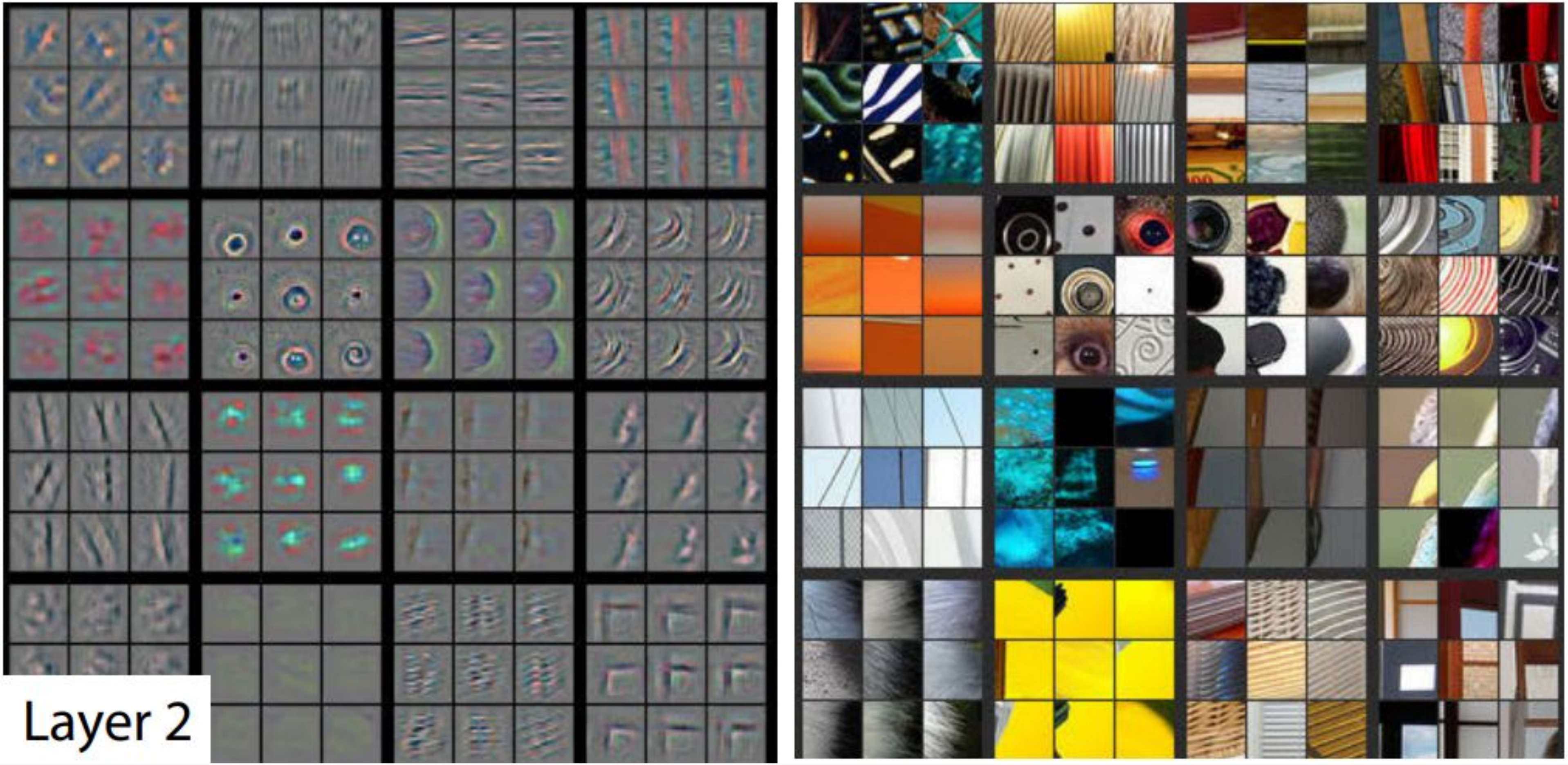
# Feature? Show me!

Low level features

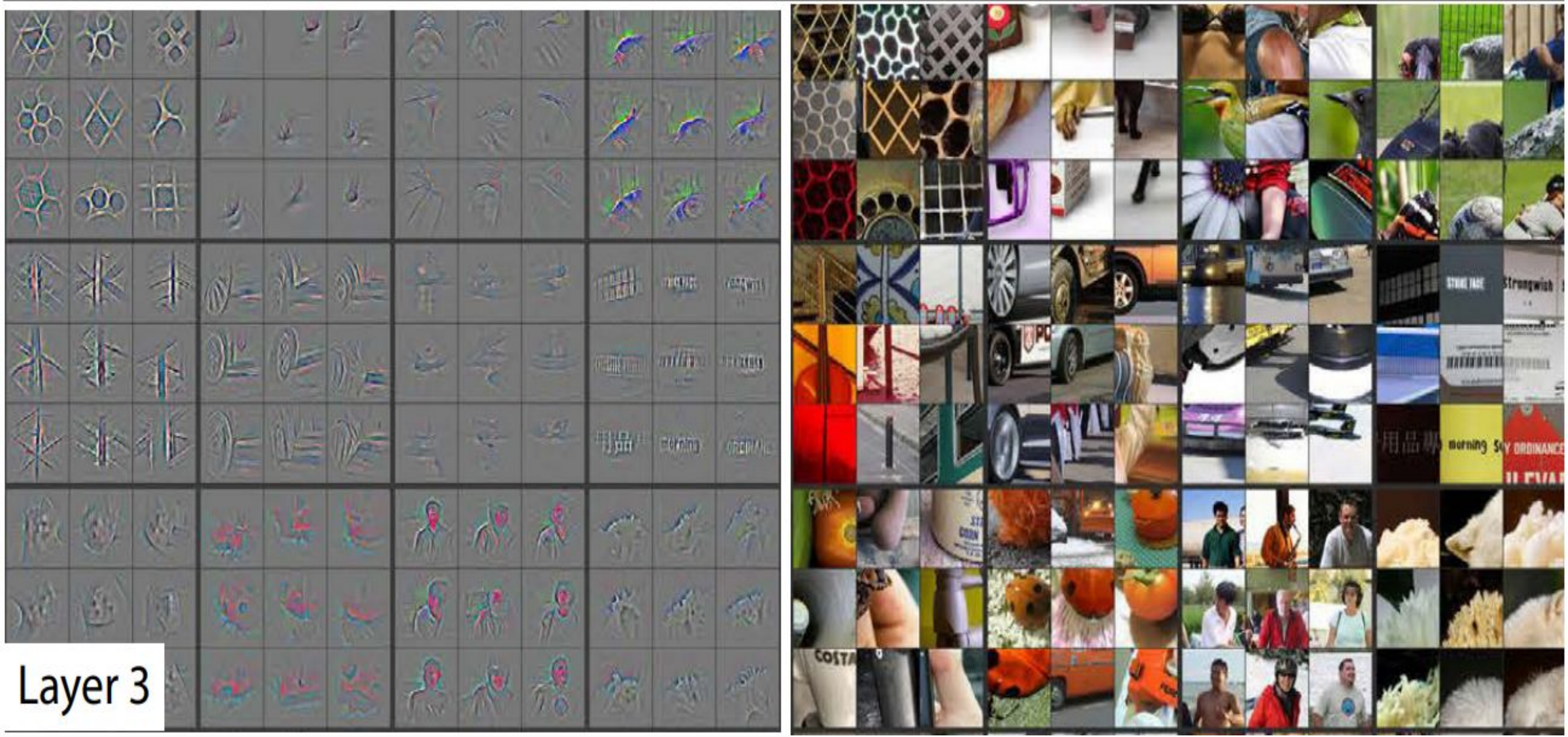
Edges, color changes, diagonals, etc.



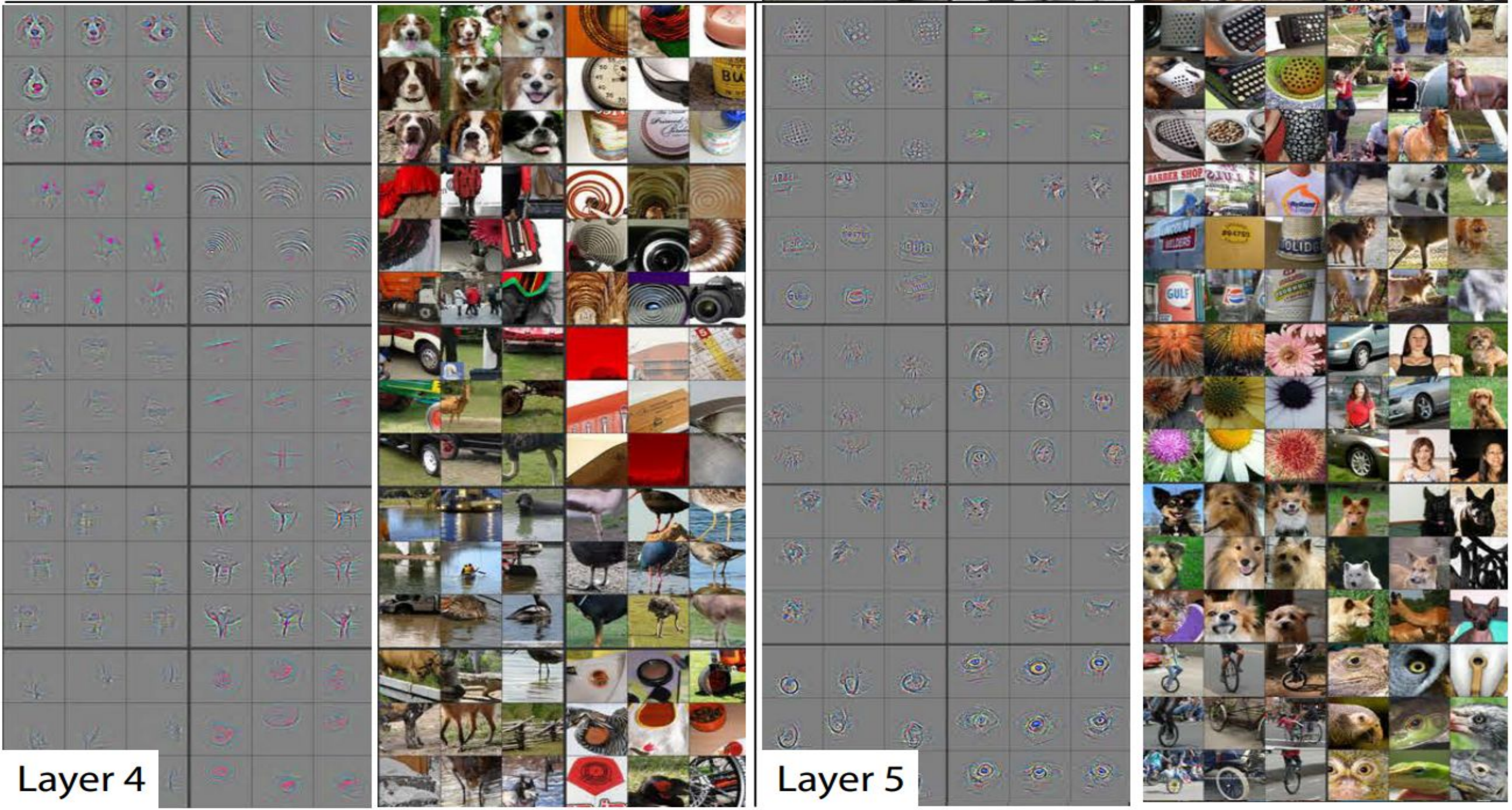

Zeiler M., Fergus R.: Visualizing and Understanding Convolutional Networks. In: ECCV (2014)




Zeiler M., Fergus R.: Visualizing and Understanding Convolutional Networks. In: ECCV (2014)



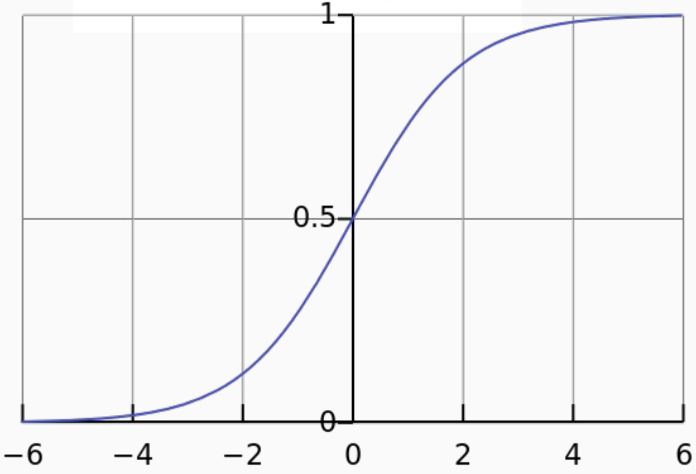

Zeiler M., Fergus R.: Visualizing and Understanding Convolutional Networks. In: ECCV (2014)




Zeiler M., Fergus R.: Visualizing and Understanding Convolutional Networks. In: ECCV (2014)

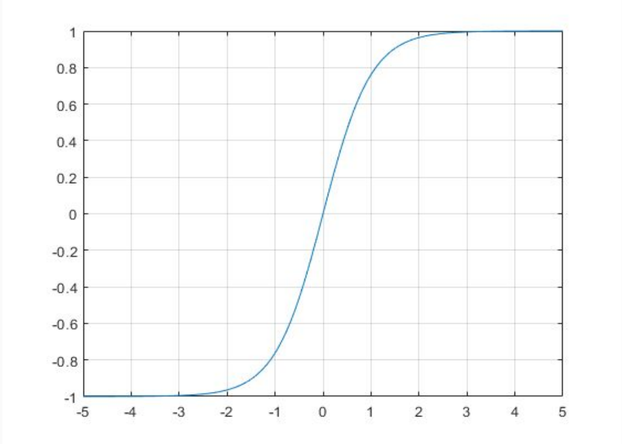
# Activation Functions

$$A = \frac{1}{1+e^{-x}}$$



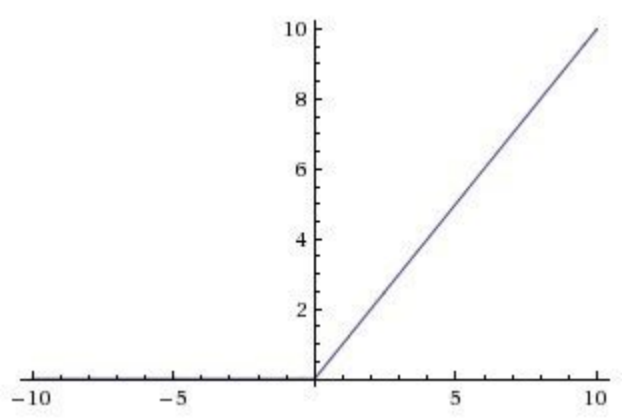

Sigmoid Function

$$f(x) = tanh(x) = \frac{2}{1+e^{-2x}} - 1$$

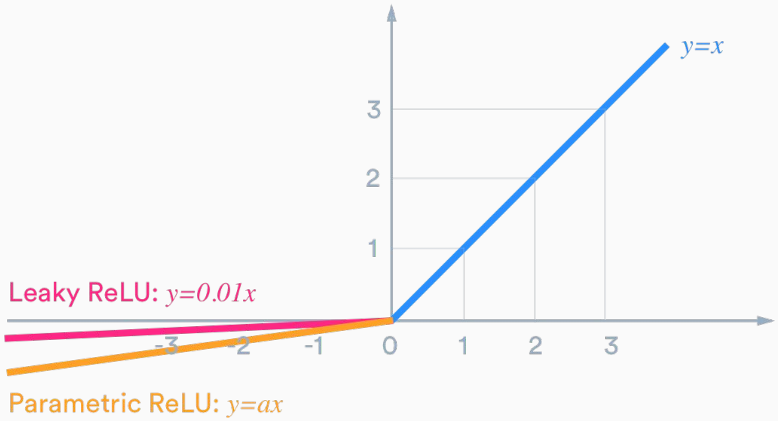



tanh Function

# Activation Functions




ReLU Function




Leaky ReLU Functions

# Activation Functions

Helps the function to make itself nonlinear

Since the function we use is the simple linear equation, we need to change it and make it non-linear prediction function.

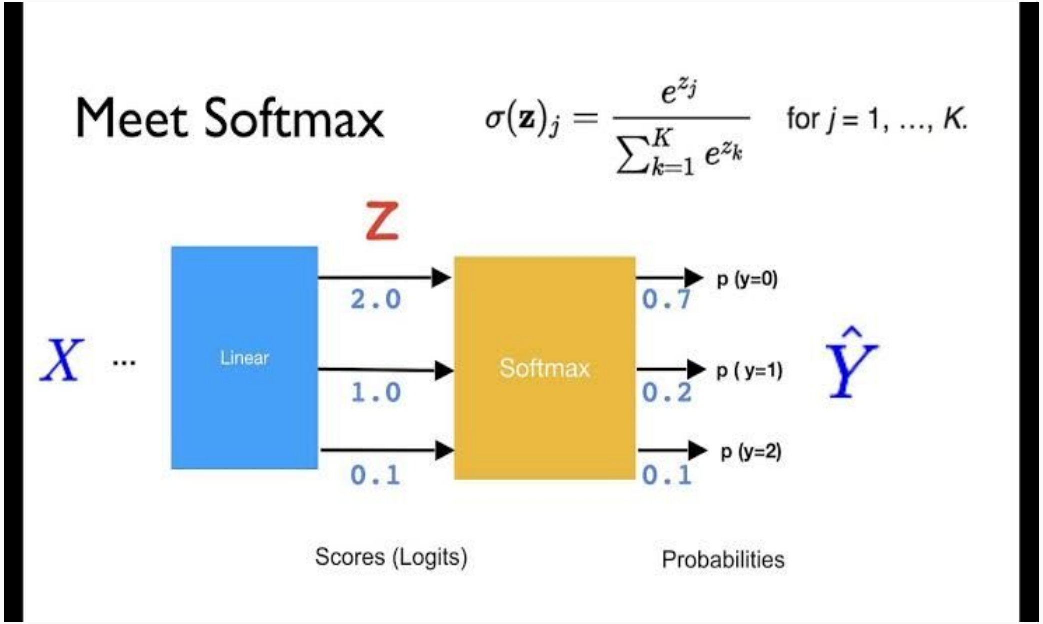
Other Activation Functions: https://en.wikipedia.org/wiki/Activation_function

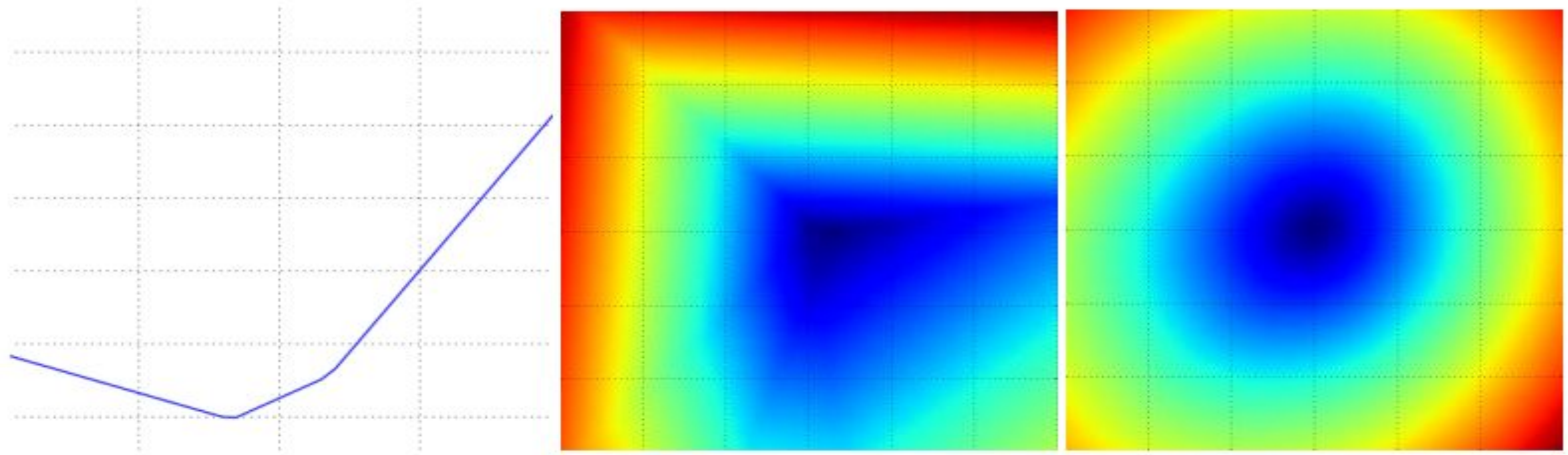
# Loss Function

Binary loss function was to classify only two different classes (dog, cat), (malignant, safe)

Since we have $K \geq 2$, we need to use something different

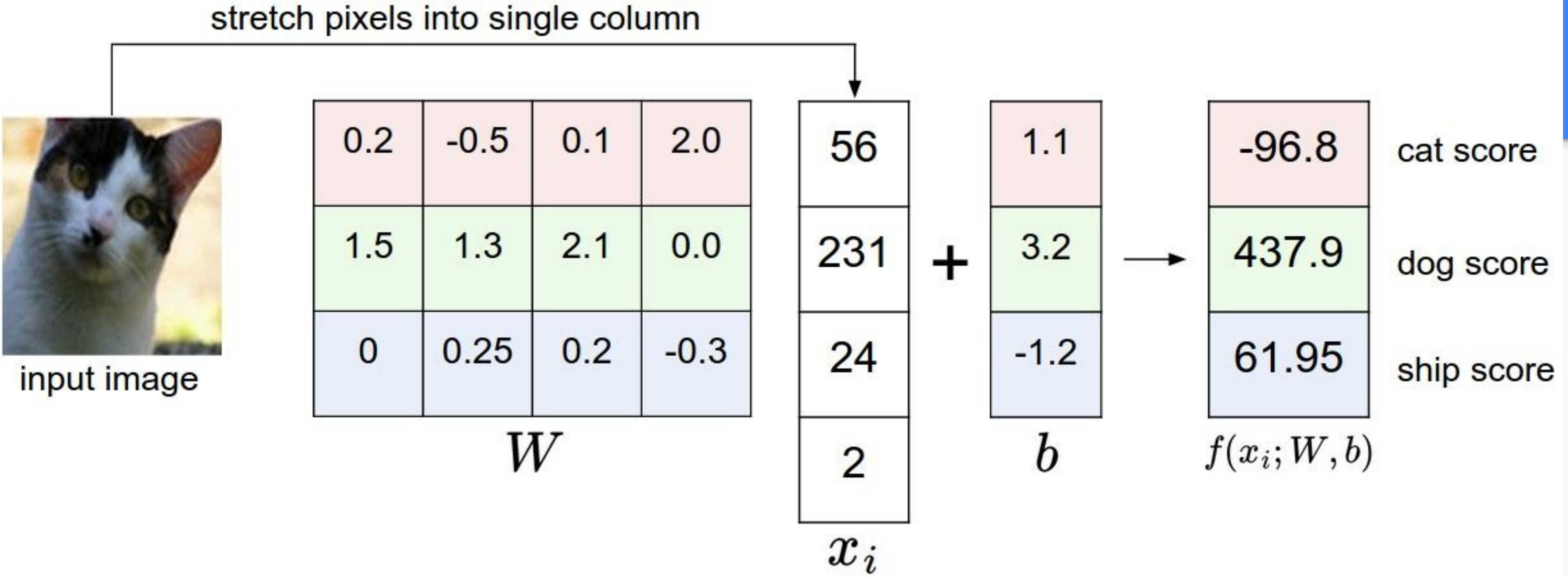
# Softmax Classifier



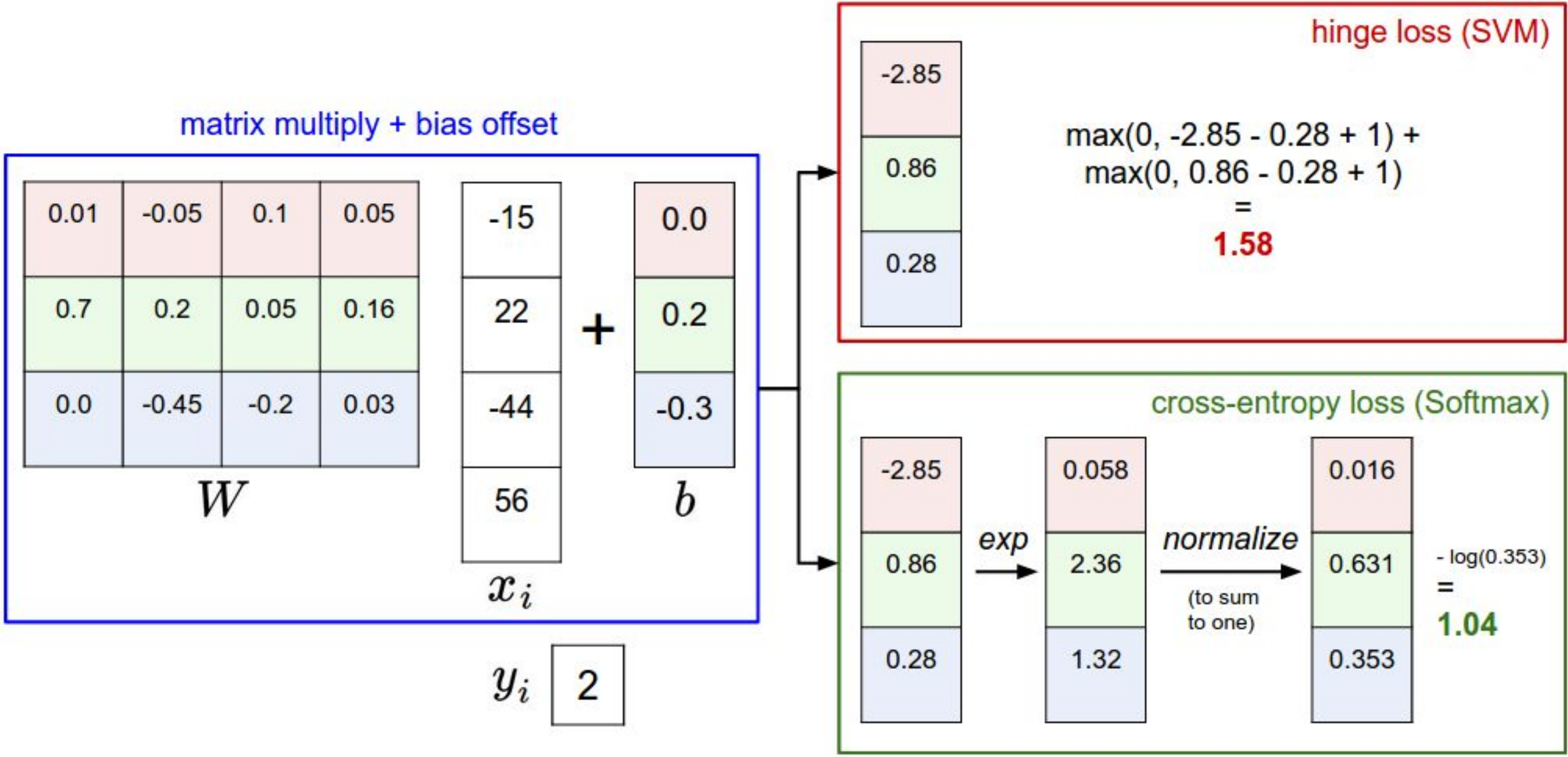


Loss function landscape for the Multiclass SVM (without regularization) for one single example (left,middle) and for a hundred examples (right) in CIFAR-10. Left: one-dimensional loss by only varying **a**. Middle, Right: two-dimensional loss slice, Blue = low loss, Red = high loss. Notice the piecewise-linear structure of the loss function. The losses for multiple examples are combined with average, so the bowl shape on the right is the average of many piece-wise linear bowls (such as the one in the middle).

Stanford CS231n Optimization Course Note




Stanford CS231n Optimization Course Note




Stanford CS231n Optimization Course Note