SC201 Lecture 13

Optimization (_____)

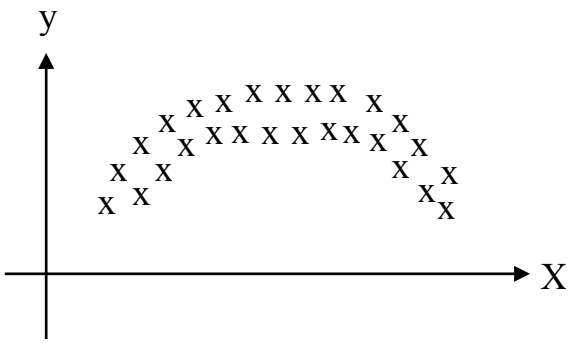
< _____ (Moving Averages)>

$V_t = \beta V_{t-1} + (1-\beta)X_t$ _____

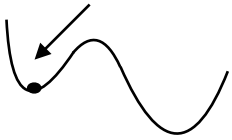
_____ = _____ _____ = _____

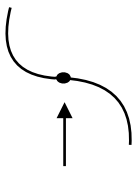
_____ = _____ _____ = _____

_____ = _____ _____ = _____



_____





_____

(a) SGD with _____

$W = W - \alpha \frac{dL}{dW}$ # _____

→ $W = W - \alpha$_____ where _____

設 $\beta = 0.9$ (1) $W = W - \alpha$ _____

(2) $W = W - \alpha$ _____

⋮

(h) $W = W - \alpha$ _____

(b) SGD with _____

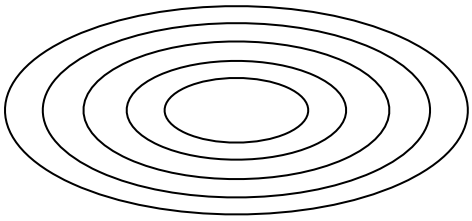
→ Update depend on _____

_____ = _____

$W = W - \alpha V_t$ where V_t = _____

(c) _____ (Adaptive Gradient)

- _____ the learning on _____axis
- _____ the learning on _____axis



→ 

t	$\frac{dL}{dW}$	α_{new}
1	1	
2	1	
3	5	

愈____$\frac{dL}{dW}$ 愈____

→ $\alpha_{new}^{陡}$ 愈 _____

(d) _____& _____

Adagrad learning rate may be close to _____

→ _____ where Vt = _____

(e) _____ → _____ + _____

Recipe for Training NN

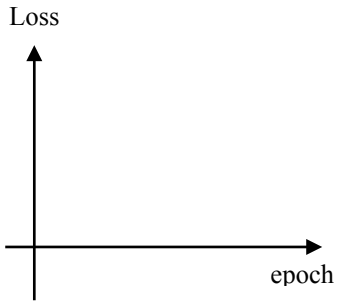
(1) Data Preprocessing
(2) Weight Initialization
(3) Normalization Layer
(4) Hyperparameter Tuning
(5) Optimization

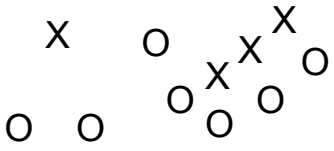
(6) _____

(Prevent from _____)

① _____

② _____



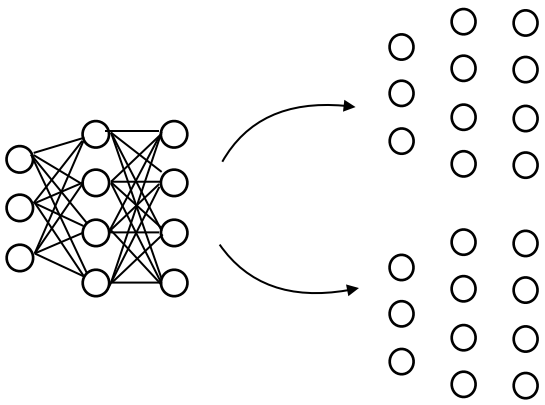


Full Loss = Loss + Loss

L2 Regularization =

L1 Regularization =

Dropout



K1 = np.dot(W1.T, X) + B1

A1 = np.maximum(0, K1)

_____

_____ = _____

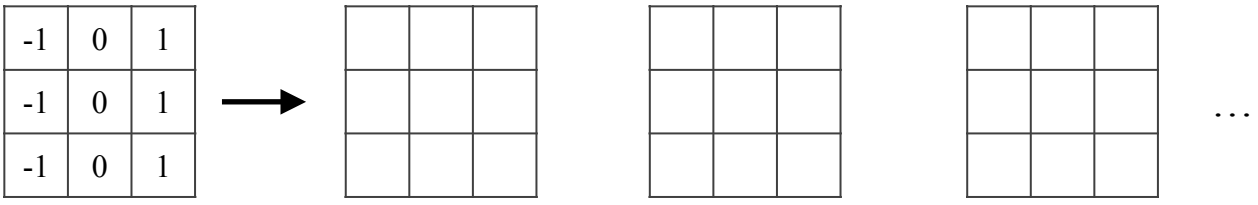
K2 = np.dot(W2.T, A1) + B2

A2 = np.maximum(0, K2)

_____

Computer Vision – _____ (CNN)

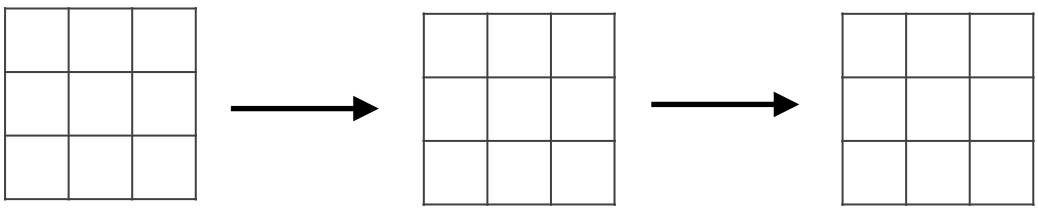
Detect Edges



_____ filter → train1

_____ filter → train2

Jerry filter → train3



#怎麼理解 2nd Conv2D?

Effective Receptive Field

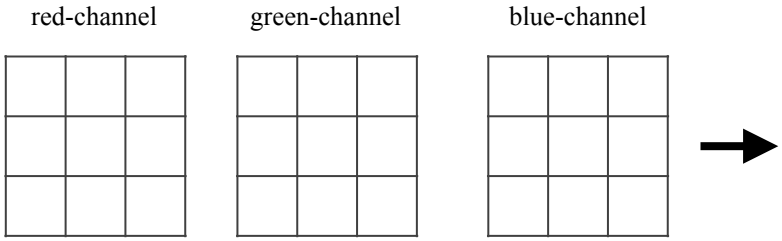
- Convolve 後的 _____ 都是原圖 _____ 的特徵

  (Covolve 愈 ____次，特徵愈_____)

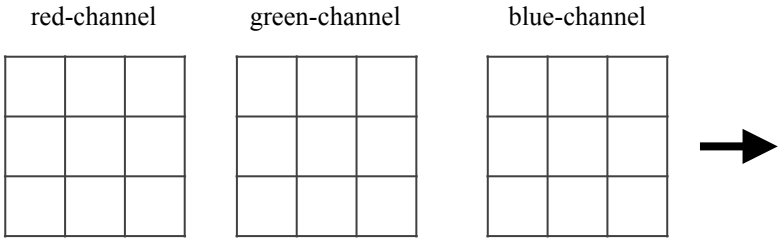
  - 一律使用 _____
  - 格子在Convole後變少，怎麼辦？ → _____

Detect edges on RGB

Kernel1: Get red-channel edges



Kernel2: Get all-channel edges



Input	Filter	Padding	Stride	Output

若使用8個3x3x3的filters, padding=1, stride=1，

對3x32x32的RGB影像convolve，output = _____