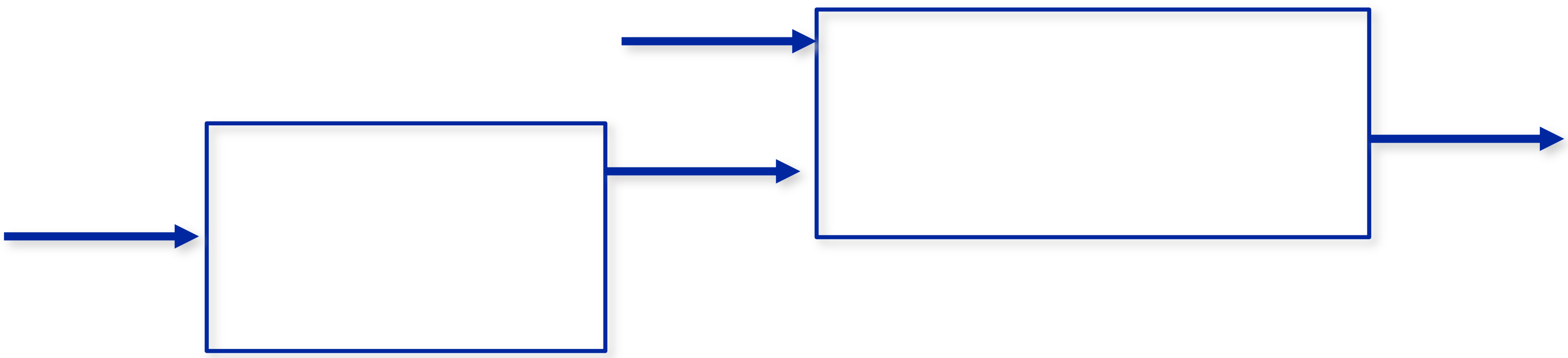
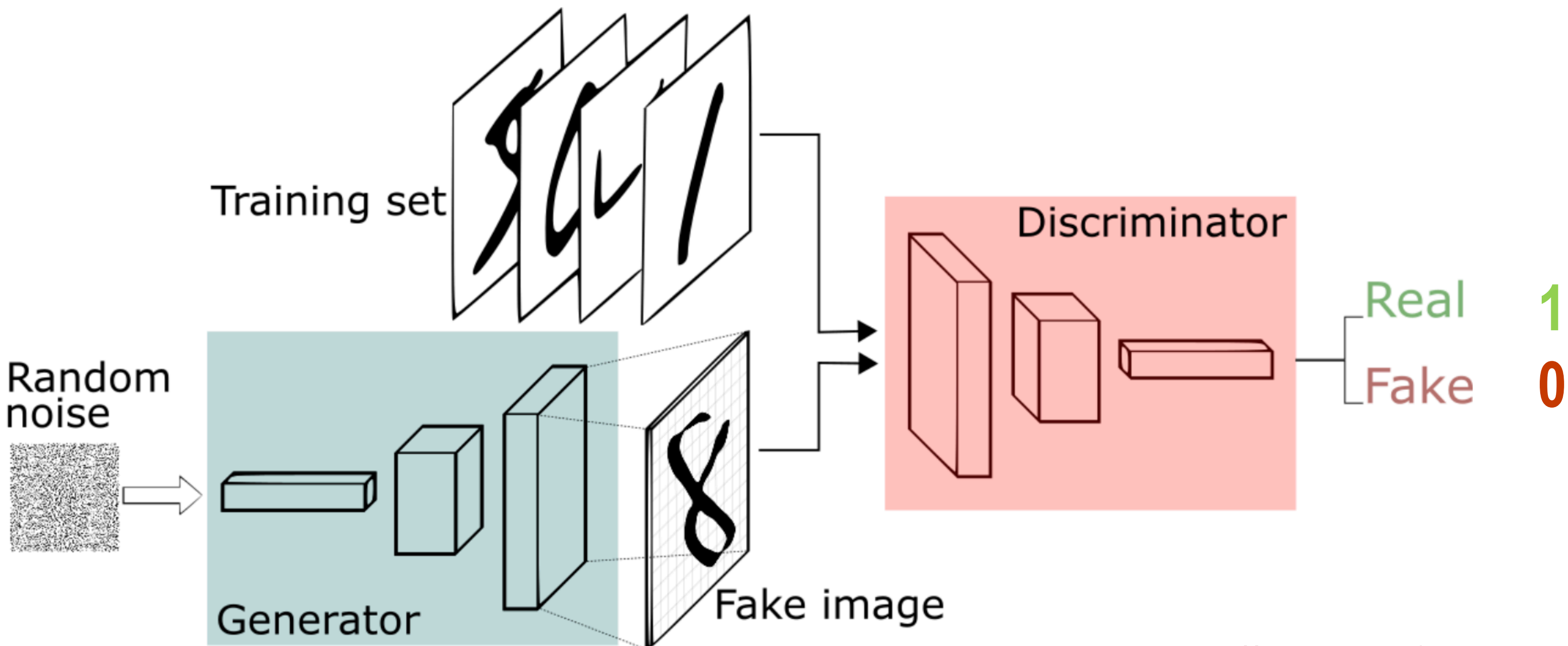
Midterm Review

2019

GANS



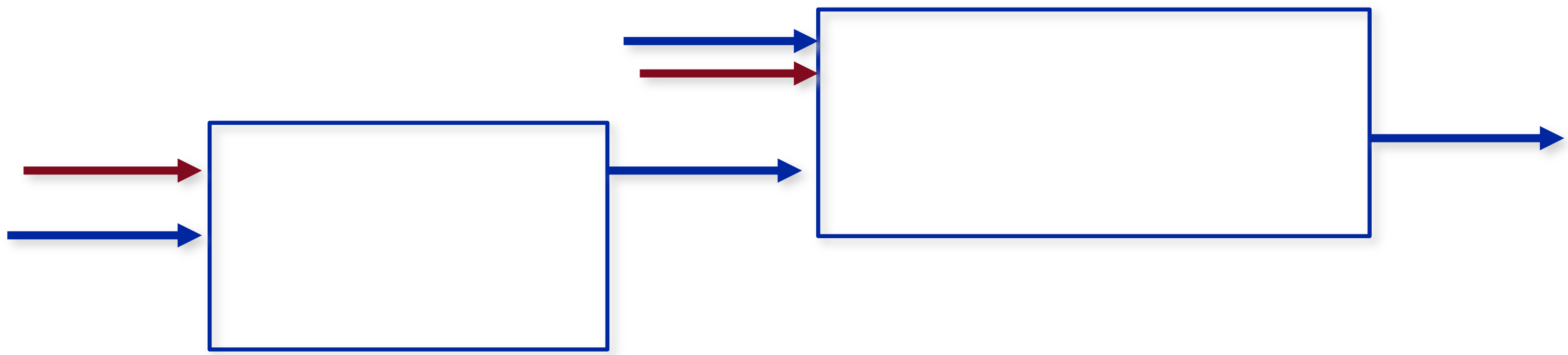
Generative Adversarial Networks: GANS



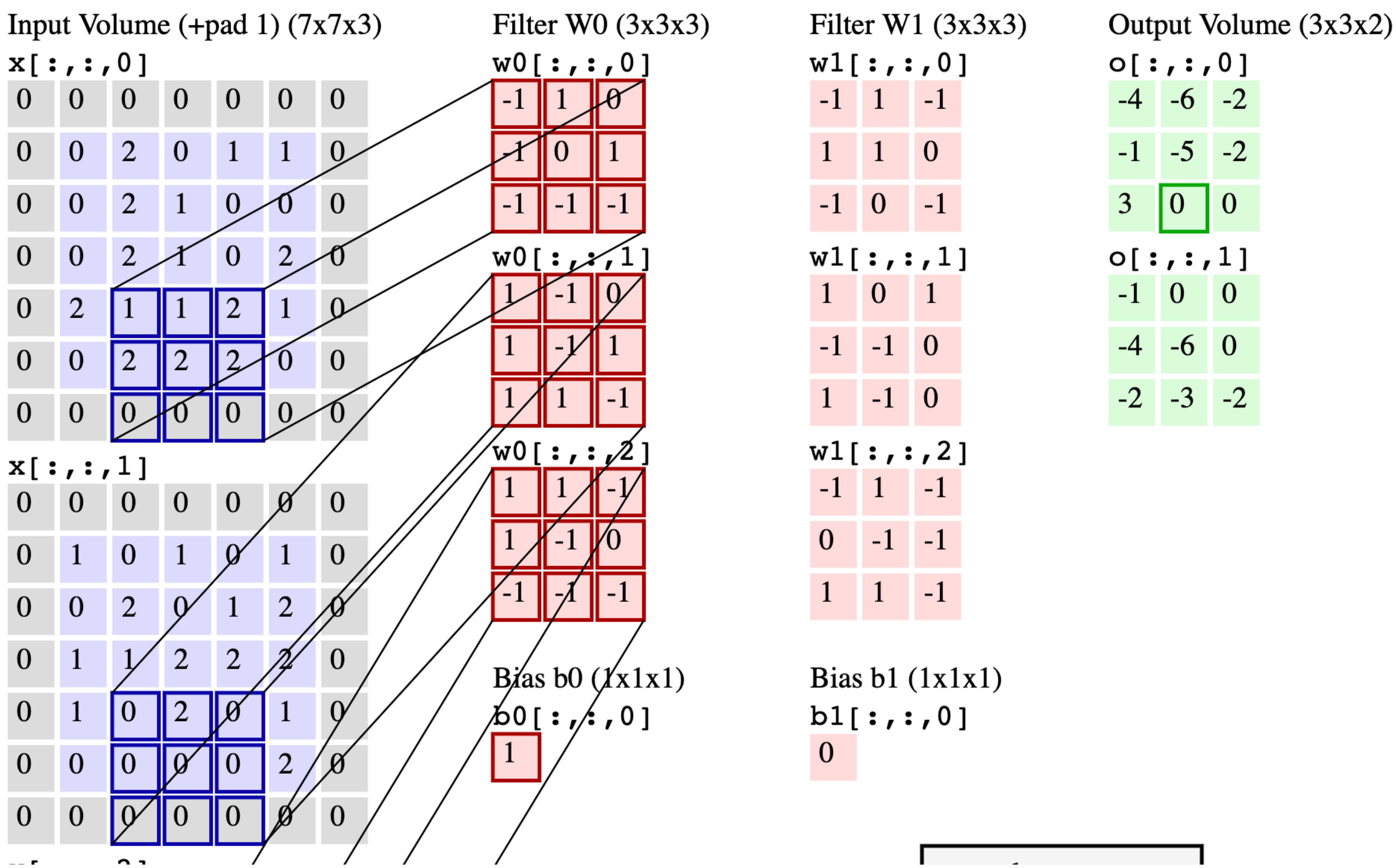
G: argmin log(1 − D(G(noise)))

https://medium.freecodecamp.org/an-intuitive-introduction-to-generative-adversarial-networks-gans-7a2264a81394

Conditional GANS



CNN


Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 2 0 1 1 0
0 0 2 1 0 0 0
0 0 2 1 0 2 0
0 2 1 1 2 1 0
0 0 2 2 2 0 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 1 0 1 0 1 0
0 0 2 0 1 2 0
0 1 1 2 2 2 0
0 1 0 2 0 1 0
0 0 0 0 0 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
-1 1 0
-1 0 1
-1 -1 -1
w0[:,:,1]
1 -1 0
1 -1 1
1 1 -1
w0[:,:,2]
1 1 -1
1 -1 0
-1 -1 -1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
-1 1 -1
1 1 0
-1 0 -1
w1[:,:,1]
1 0 1
-1 -1 0
1 -1 0
w1[:,:,2]
-1 1 -1
0 -1 -1
1 1 -1
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
-4 -6 -2
-1 -5 -2
3 0 0
o[:,:,1]
-1 0 0
-4 -6 0
-2 -3 -2

Information and friends

- Entropy of the expected value of ____
- KL divergence is the expected value of ___
- Information gain is the difference between ___

Kullback Leibler divergence

- ***P*** = true distribution;
- ***Q*** = alternative distribution that is used to encode data
- **KL** divergence is the expected extra message length per datum that must be transmitted using ***Q***

$$D_{KL}(P \,||\, Q) = \Sigma_i P(x_i) \log (P(x_i)/Q(x_i))$$

$$= -\Sigma_i P(x_i) \log Q(x_i) + \Sigma_i P(x_i) \log P(x_i)$$

$$= \quad H(P,Q) \qquad\qquad - H(P)$$

$$= \text{Cross-entropy} \qquad - \text{entropy}$$

$$D(\, p(\theta \mid x, x') \,||\, p(\theta \mid x))$$

- **Measures how different the two distributions are**

Scale invariance

- Decision tree?
- k-nn?
- OLS?
- Elastic net?
- L_0 penalized regression?
- SVM?

k-class logistic regression

$$P(Y=k|\mathbf{x},\mathbf{w}) = \frac{\exp\{\mathbf{w}_k^\top \mathbf{x}\}}{\sum_{k'=1}^{K} \exp\{\mathbf{w}_{k'}^\top \mathbf{x}\}}, \quad \text{for} \quad k = 1, \ldots, K$$

Prediction: $y = \mathrm{argmax}_k(\mathbf{w}_k^T \mathbf{x})$

Kernel functions $k(\mathbf{x}_1,\mathbf{x}_2)$

- **Measure similarity or distance?**
- **How to check if something is a kernel function?**
  - Compute a Kernel matrix with elements $k(\mathbf{x}_i,\mathbf{x}_j)$
  - Make sure its eigenvalues are non-negative
- **Example:** $k(\mathbf{x}_i,\mathbf{x}_j) = x_{i1} + x_{i2} + x_{j1} + x_{j2}$
  - Try the single point $\mathbf{x} = (1,-2)$
  - $K(x,x) = 1-2+1-2 = [-3]$ which is a matrix with eigenvalue -3