Midterm Review

2019
GANS
Generative Adversarial Networks: GANS

G: $\text{argmin} \ \log(1 - D(G(\text{noise})))$

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

<table>
<thead>
<tr>
<th>Input Volume (+pad 1) (7x7x3)</th>
<th>Filter W0 (3x3x3)</th>
<th>Filter W1 (3x3x3)</th>
<th>Output Volume (3x3x2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x[:, :, 0] )</td>
<td>( w0[:, :, 0] )</td>
<td>( w1[:, :, 0] )</td>
<td>( o[:, :, 0] )</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>-1 1 -1</td>
<td>-1 1 -1</td>
<td>-4 -6 -2</td>
</tr>
<tr>
<td>0 0 2 0 1 1 0 0</td>
<td>-1 0 1</td>
<td>1 1 0</td>
<td>-1 -5 -2</td>
</tr>
<tr>
<td>0 0 2 1 0 0 0 0</td>
<td>-1 -1 -1</td>
<td>-1 0 -1</td>
<td>3 0 0</td>
</tr>
<tr>
<td>0 2 1 1 2 1 0 0</td>
<td>1 -1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 2 2 2 0 0 0</td>
<td>1 -1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>1 1 -1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x[:, :, 1] )</td>
<td>( w0[:, :, 1] )</td>
<td>( w1[:, :, 1] )</td>
<td>( o[:, :, 1] )</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td>1 1 -1</td>
<td>1 0 1</td>
<td>-1 0 0</td>
</tr>
<tr>
<td>0 1 0 1 0 1 0 0</td>
<td>-1 -1 0</td>
<td>-1 -1 0</td>
<td>-4 -6 0</td>
</tr>
<tr>
<td>0 0 2 0 1 2 0 0</td>
<td>1 1 -1</td>
<td>1 -1 0</td>
<td>-2 -3 -2</td>
</tr>
<tr>
<td>0 1 1 2 2 2 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 1 0 2 0 1 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 0 0 2 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bias \( b0 (1x1x1) \)

Bias \( b1 (1x1x1) \)
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 $Q$

\[
D_{KL}(P \parallel Q) = \sum_i P(x_i) \log (P(x_i)/Q(x_i))
\]

\[
= - \sum_i P(x_i) \log Q(x_i) + \sum_i P(x_i) \log P(x_i)
\]

\[
= H(P,Q) - H(P)
\]

= Cross-entropy − entropy

\[
D( p(\theta \mid x, x') \parallel p(\theta \mid x))
\]

- 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

\[ P(Y = k | \mathbf{x}, \mathbf{w}) = \frac{\exp\{\mathbf{w}_k^T \mathbf{x}\}}{\sum_{k'=1}^{K} \exp\{\mathbf{w}_{k'}^T \mathbf{x}\}}, \quad \text{for} \quad k = 1, \ldots, K \]

Prediction: \( y = \text{argmax}_k (\mathbf{w}_k^T \mathbf{x}) \)
Kernel functions $k(x_1, x_2)$

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