Prova AAP - GANs

Thiago Malta Coutinho - 2014.123.335

1

1. Questão 1

A estratégia min-max, ou minimax, consiste em mínimizar o máximo de dano, ou custo, que o oponente pode causar no jogador que aplica a estratégia. Já a estratégia max-min visa máximizar o mínimo de custo causado ao oponente.

2. Questão 2

Se:
B escolhe B1, A escolhe A3;
B escolhe B2, A escolhe A1;
B escolhe B3, A escolhe A3;
B escolhe B4, A escolhe A1;

Não há estratégia dominante para A, existe 50% de chance de ocorrência para escolha para cada um dos casos A3 e A1. Mas como a estratégia é max-min, ao somar o ganho/perda total obtemos:

B escolhe A1: a soma dos payoff é igual à 12.
B escolhe A3: a soma dos payoff é igual à -2.
Portanto a estratégia que máximiza o payoff mínimo de B é A escolher A3.

3. Questão 3

Porque a função de perda do gerador é o negativo da função de perda do discriminador(próximo disso), e vice-versa. Ou seja, o que o gerador ganha, o discriminador perde; e o que o discriminador perde, o gerador ganha. Esse tipo de jogo é chamado de jogo de soma-zero, também conhecido como minimax. O jogo minimax visa minimizar o máximo de perda que o oponente pode causar. Ao analisar as funções de custo, como uma é o negativo da outra, minimizar o máximo da perda causada é minimizar o custo dado todas as estratégias possíveis que o oponente possa adotar.

A função V(G,D) é denotada por:

$$V(G, D) = V(\theta^{(D)}, \theta^{(G)}) = \frac{1}{2}\mathbb{E}_{p_{data}}[log(D(x)] + \frac{1}{2}\mathbb{E}_z[log(1 - D(G(z)))] \tag{1}$$

A estratégia para G é:

$$min_{\theta(G)} max_{\theta(D)} V(\theta^{(D)}, \theta^{(G)}) \tag{2}$$

E para D é:

$$min_{\theta(D)} max_{\theta(G)} V(\theta^{(D)}, \theta^{(G)}) \tag{3}$$

4. Questão 4

No início do treinamento da rede, o discriminador consegue discriminar bem as saídas do gerador, atribuindo valores baixos de probabilidades as saídas, porque o gerador ainda não aprendeu o suficiente para gerar exemplos minimamente bons.
A função de perda da Equação 4 tem vetor gradiente de magnitude pequena quando a probabilidade fornecida por D(z) é pequena. Portanto ao utilizar a Equação 4, os vetores de pesos da rede geradora não serão atualizados de maneira significativa, levando muito tempo para treinar ou até impossibilitando solução para o problema.

$$J^{(G)} = \mathbb{E}_z log(1 - D(G(z)) \tag{4}$$

Já a Equação 5 possui vetor gradiente com magnitude grande para baixas probabilidades e magnitude pequena para probabilidades grandes. Esse comportamento é bastante desejável para o problema, pois quando o discriminador está conseguindo discriminar bem as saídas do gerador, os pesos do gerador devem ser atualizados com maior intensidade; já quando o gerador está conseguindo gerar saídas mais difíceis de classificar, os seus pesos devem ser atualizados de forma mais suave, preservando o aprendizado da rede.

$$J^{(G)} = -\mathbb{E}_z log(D(G(z)) \tag{5}$$

5. Questão 5

5.1. a)

O discriminador ótimo será obtido quando o gerador for ótimo. Um gerador ótimo gera amostras com distribuição de probabilidade igual à distribuição original dos dados:

$$p_{data} = p_{model}$$

Como o gerador ótimo gera amostras indistinguíveis da distribuição original, o discriminador ótimo irá discriminar com probabilidade igual à:

$$D(x)_{otimo} = \frac{1}{2}$$

5.2. b)

Como $D(x)_{otimo} = \frac{1}{2}$, a partir da função de custo J, calcula-se o valor:

$$J = \frac{1}{2}\mathbb{E}_{pdata} log(D(x)) + \frac{1}{2}\mathbb{E}_z log(1 - D(G(Z))) = \frac{1}{2}\mathbb{E}_{pdata} log(\frac{1}{2}) + \frac{1}{2}\mathbb{E}_z log(1 - \frac{1}{2}))$$

$$J = log(\frac{1}{2})$$