1. Выберите какая из этих функций не является выпукло-вогнутой (выпуклой по $x \in \mathbb{R}^n$ и вогнутой по $y \in \mathbb{R}^n$) в независимости от условий:

(a) $\|x\|_2^2 - \|y\|_2^2$;

(b) $x^\top y$;

(c) $f(x) + yg(x)$, где $y \in \mathbb{R}$, и f, g - произвольные выпуклые функции;

(d) $f(x) + x^\top Ay - g(y)$, где f, g - произвольные выпуклые функции, а A - произвольная матрица.

(a) Очевидно, что по x функция выпуклая, а по y вогнутая

(b) Относительно переменной x функция является линейной (а линейная функция и выпуклая и вогнутая одновременно), тоже самое относительно переменной y. Значит функция выпукло-вогнутая

(c) Пусть $f(x) = g(x) = \|x\|_2^2$. Тогда относительно переменной x при $y = -2$ функция будет невыпуклая. Следовательно эта функция не является выпукло-вогнутой

(d) Относительно переменной x функция является выпуклой (как линейная комбинация выпуклой функции $f(x)$ и линейной функции $x^\top Ay - g(y)$). Также относительно переменной y функция является вогнутой (как линейная комбинация линейной функции $f(x) + x^\top Ay$ и вогнутой функции $-g(y)$)

Ответ: c

2. Рассмотрим следующую седловую задачу:

$$\min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} f(x, y) := xy$$

Седловой точкой (решением) этой задачи будет точка:

(a) $x^* = -1, y^* = -1$;

(b) $x^* = 0, y^* = 0$;

(c) $x^* = 1, y^* = -1$;

(d) $x^* = -1, y^* = 1$.

Ответ: b

3. Для решения седловой задачи из предыдущего пункта можно воспользоваться градиентным спуском-подъемом с некоторым постоянным шагом $\gamma > 0$ и стартовой точкой $(x_0, y_0) = (1, 1)$:

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \gamma \begin{pmatrix} \nabla_x f(x, y) \\ -\nabla_y f(x, y) \end{pmatrix}$$

Что можно сказать о расстоянии до решения $\Psi_k := (x_k - x^*)^2 + (y_k - y^*)^2$:

(a) $\Psi_2 < \Psi_1 < \Psi_0$ при достаточно малом γ;

(b) $\Psi_1 < \Psi_0$ при достаточно малом γ, но $\Psi_2 > \Psi_1$ при любом γ;

(c) $\Psi_1 > \Psi_0$ при любом γ, но $\Psi_2 < \Psi_1$ при достаточно малом γ;

(d) $\Psi_2 > \Psi_1 > \Psi_0$ при любом γ.

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} - \gamma \begin{pmatrix} y_k \\ -x_k \end{pmatrix} = \begin{pmatrix} x_k - \gamma y_k \\ y_k + \gamma x_k \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 - \gamma y_0 \\ y_0 + \gamma x_0 \end{pmatrix} = \begin{pmatrix} 1-\gamma \\ 1+\gamma \end{pmatrix}$$

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 - \gamma y_1 \\ y_1 + \gamma x_1 \end{pmatrix} = \begin{pmatrix} 1-\gamma-\gamma(1+\gamma) \\ 1+\gamma+\gamma(1-\gamma) \end{pmatrix} = \begin{pmatrix} 2-(\gamma+1)^2 \\ 2-(\gamma-1)^2 \end{pmatrix}$$

$$\Psi_0 = (x_0 - x^*)^2 + (y_k - y^*)^2 = (1-0)^2 + (1-0)^2 = 2$$

$$\Psi_1 = (x_1 - x^*)^2 + (y_1 - y^*)^2 = (1-\gamma)^2 + (1+\gamma)^2 = 2(\gamma^2+1)$$

$$\Psi_2 = (x_2 - x^*)^2 + (y_2 - y^*)^2 = (2-(\gamma+1)^2)^2 + (2-(\gamma-1)^2)^2 = 2(\gamma^2+1)^2$$

$$\gamma > 0 \Rightarrow \Psi_2 > \Psi_1 > \Psi_0$$

Ответ: d

4. Для решения седловой задачи из 2го пункта можно воспользоваться экстра градиентом (экстра шагом) с некоторым шагом $\gamma > 0$ и стартовой точкой $(x_0, y_0) = (1, 1)$. Что можно сказать о расстоянии до решения $\Psi_k := (x_k - x^*)^2 + (y_k - y^*)^2$:

(a) $\Psi_2 < \Psi_1 < \Psi_0$ при достаточно малом γ;
(b) $\Psi_1 < \Psi_0$ при достаточно малом γ, но $\Psi_2 > \Psi_1$ при любом γ;
(c) $\Psi_1 > \Psi_0$ при любом γ, но $\Psi_2 < \Psi_1$ при достаточно малом γ;
(d) $\Psi_2 > \Psi_1 > \Psi_0$ при любом γ.

$$\begin{pmatrix} x_{1/2} \\ y_{1/2} \end{pmatrix} = \begin{pmatrix} x_0 - \gamma y_0 \\ y_0 + \gamma x_0 \end{pmatrix} = \begin{pmatrix} 1-\gamma \\ 1+\gamma \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 - \gamma y_{1/2} \\ y_0 + \gamma x_{1/2} \end{pmatrix} = \begin{pmatrix} 1-\gamma(1+\gamma) \\ 1+\gamma(1-\gamma) \end{pmatrix} = \begin{pmatrix} -\gamma^2-\gamma+1 \\ -\gamma^2+\gamma+1 \end{pmatrix}$$

$$\begin{pmatrix} x_{3/2} \\ y_{3/2} \end{pmatrix} = \begin{pmatrix} x_1 - \gamma y_1 \\ y_1 + \gamma x_1 \end{pmatrix} = \begin{pmatrix} -\gamma^2-\gamma+1-\gamma(-\gamma^2+\gamma+1) \\ -\gamma^2+\gamma+1+\gamma(-\gamma^2-\gamma+1) \end{pmatrix} = \begin{pmatrix} (1+\gamma)(\gamma^2-3\gamma+1) \\ (1-\gamma)(\gamma^2+3\gamma+1) \end{pmatrix}$$

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 - \gamma y_{3/2} \\ y_1 + \gamma x_{3/2} \end{pmatrix} = \begin{pmatrix} -\gamma^2-\gamma+1-\gamma(1-\gamma)(\gamma^2+3\gamma+1) \\ -\gamma^2+\gamma+1+\gamma(1+\gamma)(\gamma^2-3\gamma+1) \end{pmatrix} = \begin{pmatrix} \gamma^4+2\gamma^3-3\gamma^2-2\gamma+1 \\ \gamma^4-2\gamma^3-3\gamma^2+2\gamma+1 \end{pmatrix}$$

$$\Psi_0 = 2$$

$$\Psi_1 = x_1^2 + y_1^2 = 2(\gamma^4 - \gamma^2 + 1)$$

$$\Psi_2 = x_2^2 + y_2^2 = 2(\gamma^8 - 2\gamma^6 + 3\gamma^4 - 2\gamma^2 + 1)$$

$$\gamma \in (0;1) \Rightarrow \Psi_2 < \Psi_1 < \Psi_0$$

Ответ: a

5. Выберите верную формулировку классической задачи обучения GAN:
(a) $\max_G \min_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log D(G(z))]$;
(b) $\max_G \min_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]$;
(c) $\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log D(G(z))]$;
(d) $\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]$.

Здесь G - модель генератора, D - модель дискриминатора (выдает число от 0 до 1 - вероятность того, что данные на входе настоящие), p_{data} - распределение реальных данных (картинок), p_z - распределение из которого генератор создает "фейковые"данные

Ответ: d

Ответы:

1. c
2. b
3. d
4. a
5. d