سوال ۱

الف) غلط، زیرا ممکن است سایر ویژگی ها به این ویژگی حساس وابسته باشند، لذا نمی توان مطمئن بود که خروجی کاملاً مستقل از این ویژگی است.

ب) غلط، زیرا معمولاً روش های Data Augmentation یک مقدار باعث افزایش خطای بایاس می شوند. ولی معمولاً خطای واریانس کاهش می یابد.

ج) غلط، زیرا توزیع توام را به هر دو صورت زیر می توان نوشت و راهی برای تمایز میان آن ها نیست.

$$P(x,y) = P(x)\, P(y|x) \qquad x \to y$$

$$P(x,y) = P(y)\, P(x|y) \qquad y \to x$$

سوال ۲

باید یک تابع هدفی را تعریف کنیم که هم خطای مدل کم باشد و هم نویز مستقل از ورودی تابع f و g باشد.

یک راه حل:

$$L_1 = \min_{g} \sum_{i=1}^{n} \left(y_i - g(x_i)\right)^2 + \underbrace{I\left(x \,;\, y - g(x)\right)}_{\text{mutual Information}}$$

$$L_2 = \min_{f} \sum_{i=1}^{n} \left(x_i - f(y_i)\right)^2 + I\left(y \,;\, x - f(y)\right)$$

اگر $L_1 < L_2 \Leftarrow x \to y$

اگر $L_2 < L_1 \Leftarrow y \to x$

سوال ۵

$$\left(g(x) - g(x')\right)^2 = \left(\sum_{i=1}^{K} w_i f_i(x) - \sum_{i=1}^{K} w_i f_i(x')\right)^2$$

$$= \left(\sum_{i=1}^{K} w_i \left(f_i(x) - f_i(x')\right)\right)^2$$

$$\leq \left(\sum_{i=1}^{K} w_i^2\right) \left(\sum_{i=1}^{K} \left(f_i(x) - f_i(x')\right)^2\right)$$

$$\le \underbrace{\left(\sum_{i=1}^{k} w_i^2\right)}_{\le 1} L K \|x - x'\|_2^2$$

$$\le L K \|x - x'\|_2^2 \implies g(x) \text{ is } KL\text{-Lipschitz}$$

از آن جایی که $w_i > 0$ و $\sum w_i = 1$ است، $0 \le w_i \le 1$ ولذا: $\sum w_i^2 \le \sum w_i = 1$

راه حل دوم:

می دانیم واریانس هر متغیر تصادفی بزرگتر یا مساوی صفر است، پس داریم:

$$\mathrm{var}(x) = E(x^2) - E^2(x) \ge 0 \implies \boxed{E^2(x) \le E[x^2]} \quad (1)$$

$$(g(x) - g(x'))^2 = \Big(\underbrace{\sum_{i=1}^{K} w_i \left(f_i(x) - f_i(x')\right)}_{E[f_i(x) - f_i(x')]}\Big)^2 \overset{(1)}{\le} \underbrace{\sum_{i=1}^{n} w_i \left(f_i(x) - f_i(x')\right)^2}_{E[(f_i(x) - f_i(x'))^2]}$$

$$\le \sum_{i=1}^{n} w_i L \|x - x'\|_2^2 = L \|x - x'\|_2^2 \underbrace{\sum_{i=1}^{n} w_i}_{1} = L \|x - x'\|_2^2$$

$$\implies g(x) \text{ is } L\text{-Lipschitz} \quad (t = L)$$

راه حل سوم:

$$(g(x) - g(x'))^2 = \left(\sum_{i=1}^{n} w_i \left(f_i(x) - f_i(x')\right)\right)^2 \le \left(\sum_{i=1}^{n} w_i \left|f_i(x) - f_i(x')\right|\right)^2$$

$$\le \Big(\sum_{i=1}^{n} w_i \underbrace{\max_j \left|f_j(x) - f_j(x')\right|}_{\text{این قسمت مستقل از } i \text{ است}}\Big)^2 = \Big(\max_j \left|f_j(x) - f_j(x')\right| \underbrace{\sum_{i=1}^{n} w_i}_{1}\Big)^2$$

$$\le \left(\sqrt{L}\, \|x - x'\|_2\right)^2 = L \|x - x'\|_2^2 \implies g(x) \text{ is } L\text{-Lipschitz} \quad (L = t)$$

سوال ۲

الف) privacy این طبقه بند اصلاً خوب نیست. به طور مثال اگر یک ورودی $\underline{x}^*$ به آن بدهیم و خروجی 1 برگرداند، متوجه می شویم که $\underline{x}^*$ داخل مجموعه آموزشی وجود داشته است.

ب) راهکارهای ممکن:

۱) داده های آموزشی را نویزی کنیم

۲) از KNN با $K>1$ استفاده کنیم

۳) نمونه x_{nn} را روی یک زیربخشی از کل داده های آموزشی پیدا کنیم.

۴) ...