سوال ۱

a) غلط، خطای بایاس ناشی از انتخاب مدل اشتباه (ساده) است و با افزایش داده نمی‌توان آن را کاهش داد.

b) درست، افزودن منظم‌ساز باعث ساده‌تر شدن مدل می‌شود و لذا خطای بایاس ~~کاهش~~ افزایش می‌یابد. مدل های ساده‌تر معمولاً واریانس کمتری دارند، بنابراین خطای واریانس کاهش می‌یابد.

c) غلط، الگوریتم ID3 در واقع یک الگوریتم حریصانه است و لزوماً درخت بهینه را نمی‌یابد.

d) غلط، در حالت کلی تفاضل دو تابع کرنل، یک کرنل معتبر نیست و می‌توان نشان داد که قضیه Mercer برای آن برقرار نیست

e) درست، زیرا:

$$\phi(x) = \begin{bmatrix} x_1^2 \\ \sin x_2 x_3 \end{bmatrix} \Rightarrow g(x) = \omega^T \phi(x)$$

$$\omega = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \nearrow$$

f) غلط، در صورتی به توزیع اصلی همگرا می‌شود که $V_n \to 0$ و $nV_n \to \infty$

g) درست، زیرا روش SVM فقط به بردارهای پشتیبان (support vectors) حساس است و اهمیتی ندارد که از هر کلاس چه تعداد نمونه وجود دارد.

سوال ۲

الف)

$$L(\lambda) = \log P(D|\lambda) = \sum_{i=1}^{n} \log P(x_i|\lambda) = \sum_{i=1}^{n} \log\left(\lambda e^{-\lambda x_i}\right)$$

$$= \sum_{i=1}^{n} \log\lambda - \lambda x_i = n\log\lambda - \lambda\sum_{i=1}^{n} x_i$$

$$\frac{dL(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \boxed{\hat{\lambda}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}}$$

ب)

$$P(\lambda|D) \propto P(D|\lambda)P(\lambda) = \prod_{i=1}^{n} P(x_i|\lambda)P(\lambda)$$

$$= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \quad c\lambda^{\alpha-1} e^{-\beta\lambda} \quad = c\lambda^{n} e^{-\lambda \sum_{i=1}^{n} x_i} \quad e^{-\beta\lambda} \quad \lambda^{\alpha-1}$$

$$= c\,\lambda^{n+\alpha-1} \quad e^{-\lambda\left(\beta+\sum_{i=1}^{n} x_i\right)} \quad = \mathrm{Gamma}\Big(\lambda \,\Big|\, \underbrace{n+\alpha}_{\alpha_{new}},\ \underbrace{\beta+\sum_{i=1}^{n} x_i}_{\beta_{new}}\Big)$$

ج) بله، همانطور که در قسمت قبل محاسبه شد، توزیع احتمال پسین از نوع توزیع احتمال پیشین شده است.

د)

$$\hat{\lambda}_{MAP} = \arg\max_{\lambda} P(\lambda|D) = \frac{\alpha_{new}-1}{\beta_{new}} = \frac{n+\alpha}{\beta+\sum_{i=1}^{n} x_i}$$

هـ) بله، زیرا

$$\lim_{n\to\infty} \hat{\lambda}_{MAP} = \lim_{n\to\infty} \frac{n+\alpha}{\beta+\sum_{i=1}^{n} x_i} = \lim_{n\to\infty} \frac{n}{\sum_{i=1}^{n} x_i} = \lim_{n\to\infty} \hat{\lambda}_{ML}$$

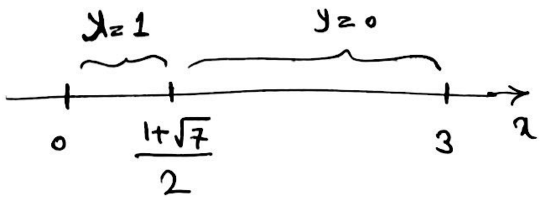
سوال ۳

در ناحیه‌ی مربوط به کلاس ۰ داریم:

$$\cancel{P(x|y=0)} \quad P(y=0|x) \geq P(y=1|x) \Rightarrow P(x|y=0)\cancel{P(y=0)} \geq P(y=1|x)\cancel{P(y=1)}$$

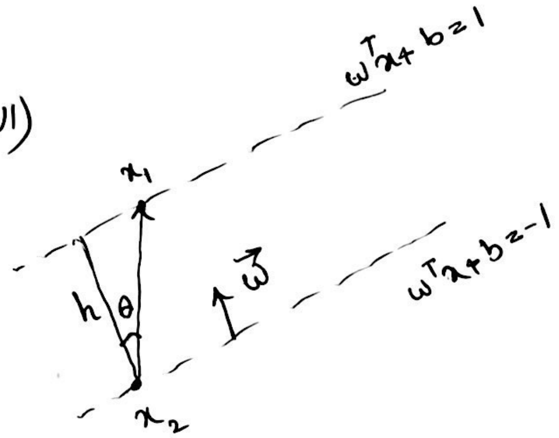
$$\Rightarrow P(x|y=0) \geq P(y=1|x) \Rightarrow \frac{x^2}{9} \geq \frac{x}{9} + \frac{1}{6} \Rightarrow 2x^2 - 2x - 3 \geq 0$$

$$x_1, x_2 = \frac{1 \pm \sqrt{1+6}}{2}$$



سوال ۴

الف)



$$\text{margin} = h = \|\overrightarrow{x_2x_1}\|_2 \cos\theta$$

$$= \frac{\|\overrightarrow{x_2x_1}\|_2 \, \|w\|_2 \cos\theta}{\|w\|_2}$$

$$= \frac{\langle \overrightarrow{x_2x_1}, \vec{w} \rangle}{\|w\|_2} = \frac{w^T(x_1 - x_2)}{\|w\|_2}$$

$$= \frac{\overbrace{w^T x_1 + b}^{1} - \overbrace{w^T x_2 - b}^{-1}}{\|w\|_2} = \frac{2}{\|w\|_2}$$

ب)

با استفاده از رابطه‌ی $w = \sum_{i=1}^{n} \lambda_i y_i x_i$ مقدار بردار w بدست می‌آید. سپس با استفاده از یکی از بردارهای پشتیبان (نمونه‌هایی که روی خط margin قرار گرفته‌اند)، مقدار b (عرض از مبدا) را محاسبه می‌کنیم:

$$w^T x_j + b = y_j \Rightarrow \boxed{b = y_j - w^T x_j}$$

که در رابطه‌ی بالا، x_j یک SV است و لذا λ_j متناظر با آن حتماً مثبت است. y_j برچسب نمونه‌ی x_j است.

ج)

مرز جداکننده:
$$g(x) = w^T \phi(x) + b = \left(\sum_{i=1}^{n} \lambda_i y_i \phi(x_i)\right)^T \phi(x) + b$$

$$= \sum_{i=1}^{n} \lambda_i y_i \underbrace{\phi^T(x_i)\,\phi(x)}_{k(x_i, x)} + b = \sum_{i=1}^{n} \lambda_i y_i \, k(x_i, x) + b$$

طبق رابطه‌ی بالا، مرز جداکننده توسط کرنل بدست می‌آید. در ترم summation فقط کافیست SV ها را وارد کنیم چون مقدار λ_i برای سایر داده‌ها صفر است.

سوال ۵

ماتریس X و بردار y را به صورت زیر تعریف می‌کنیم:

$$X = \begin{bmatrix} — & x_1^T & — \\ & \vdots & \\ — & x_n^T & — \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

لذا تابع $L(w)$ را می‌توانیم به صورت زیر بنویسیم:

$$L(w) = \|Xw - y\|_2^2 + \lambda \|w\|_2^2 = (Xw - y)^T (Xw - y) + \lambda w^T w$$

$$= w^T X^T X w - w^T X^T y - y^T X w + y^T y + \lambda w^T w$$

$$\nabla_w L(w) = 2 X^T X w - X^T y - X^T y + 2\lambda w = 0$$

$$\Rightarrow (X^T X + \lambda I) w = X^T y$$

$$\Rightarrow \boxed{w = (X^T X + \lambda I)^{-1} X^T y}$$

در رابطه‌ی بالا، I ماتریس همانی است.