سوال اول –

قسمت آ) اگر ζ_n کوچک تر از صفر باشه(منفی باشه) لذا $1-\zeta_n$ بزرگ تر از ۱ میشه و:

$$1-\zeta_n \ \geq 1 \ \rightarrow y^i(w^Tx^i+b) \geq 1-\zeta_n \ \geq 1$$

و مسئله به مسئله ماشین بردار پشتیبان با حاشیه سخت تبدیل میشه. از آن جایی که بعضی مواقع در حاشیه نرم باید margin violation رخ بده پس حتما باید بعضی مواقع $y^i(w^Tx^i+b) \geq 1$ نقض بشه لذا باید $1-\zeta_n \leq 1$ شود پس باید $\zeta_n \geq 0$ باشد.پس بی تاثیر است.

قسمت ب)

$$\mathcal{L}(w,b,\zeta,\alpha) = \frac{1}{2}w^Tw+\frac{C}{2}\sum_{i=1}^{N}{\zeta_i}^2-\sum_{i=1}^{N}[\alpha_i(y_i(w^Tx_i+b)-1+\zeta_i)]$$

$$minimize\ w.r.t\ w,b,\zeta\ and\ maximize\ w.r.t\ \alpha_i \geq 0$$

با گرفتن دوگان این مسئله بهینه سازی و سپس گرادیان ها داریم:

$$\nabla_w\mathcal{L} = w-\sum_{i=1}^{N}\alpha_iy_ix_i = 0 \ \rightarrow w^* = \sum_{i=1}^{N}\alpha_iy_ix_i$$

$$\nabla_b\mathcal{L} = \ -\sum_{i=1}^{N}\alpha_iy_i = 0$$

$$\nabla_{\zeta_i}\mathcal{L} = \frac{C}{2}\zeta_i-\alpha_i = 0 \ \rightarrow \zeta_i = \frac{2\alpha_i}{C}$$

قسمت پ)

$$\mathcal{L}(w,b,\zeta,\alpha)=\frac{1}{2}w^Tw+\frac{C}{2}\sum_{i=1}^{N}{\zeta_i}^2-\sum_{i=1}^{N}[\alpha_i(y_i(w^Tx_i+b)-1+\zeta_i)]$$

$$=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_jx_i^Tx_j+\frac{C}{2}\sum_{i=1}^{N}\left(\frac{2\alpha_i}{C}\right)^2+\sum_{i=1}^{N}\alpha_i \quad \dots$$

$$\dots \quad -\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_jx_i^Tx_j-\sum_{i=1}^{N}\alpha_iy_ib$$

$$=-\frac{1}{2}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_jx_i^Tx_j+\frac{1}{C}\sum_{i=1}^{N}(2\alpha_i)^2\right]+\sum_{i=1}^{N}\alpha_i$$

لذا مسئله دوگان برابر است با:

$$maximize_{\alpha\geq 0} \quad -\frac{1}{2}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_jy_iy_jx_i^Tx_j+\frac{1}{C}\sum_{i=1}^{N}(2\alpha_i)^2\right]+\sum_{i=1}^{N}\alpha_i$$

$$subject\ to\colon \sum_{i=1}^{N}y_i\alpha_i=0$$

قسمت ث)همانطور که در کلاس درس داشتیم، پارامتر *C* در اصل مصالحه ای بین تاثیر ***margin*** و تاثیر***violation*** است. یعنی یه جورایی میزان ***soft*** بودن مسئله.وقتی بینهایت باشد، یعنی میخواهیم مسئله به سمت ***soft margin*** برود یعنی ***margin*** کمتر نیز قابل قبول است و ***violation*** نباید رخ دهد و برعکس وقتی به سمت صفر میل میکند، یعنی مسئله به سمت ***hard margin*** بودن میرود یعنی ***margin*** زیاد میخواهیم و ***violation*** مشکلی ندارد.

سوال دوم – طبق Mercer condition باید برای هر یک تقارن نسبت به دو مولفه آن و positive semi definite بودن آن را اثبات کنیم.

۱– اگر a یک بردار در فضای n بعدی باشد و x1,x2,....,xn دلخواه را داشته باشیم:

$$a^Tk_3a=\sum_i\sum_j a_i\left(k_1(x_i,x_j)+k_2(x_i,x_j)\right)a_j$$

$$= \sum_i \sum_j a_i \left(k_1(x_i, x_j)\right) a_j + \sum_i \sum_j a_i \left(k_2(x_i, x_j)\right) a_j$$
$$= a^T k_1 a + a^T k_2 a$$

و چون هسته های k1 و k2 معتبر هستند یعنی positive semi definite هستند یعنی دو عبارت بالا بزرگ تر مساوی صفر هستند پس نتیجه شد که $a^T k_3 a$ هم بزرگ تر مساوی صفر است لذا هسته k3 هم positive semi definite است. به عنوان اثبات دیگر میتوان fi1 و fi2 را از k1 و k2 استخراج کرد و با مضرب رادیکال a در نظر گرفت(باز کردن ki بر اساس fi آن) و اثبات کرد.

حال برای متقارن بودن آن میدانیم که هسته k1 و هسته k2 ، معتبر اند یعنی نسبت به ورودی هایش متقارن هستند پس داریم:

$$k_3(x_1, x_2) = k_1(x_1, x_2) + k_2(x_1, x_2) = k_1(x_2, x_1) + k_2(x_2, x_1) = k_3(x_2, x_1)$$

لذا متقارن هم هست. پس نتیجه میشود معتبر هم هست.

۲- این قسمت نیاز به اثبات ندارد زیرا بدیهی است. از آن جا که *k1* برای خودش فضای ویژگی(*feature map*) هه *fi1* را دارد و *k2* هم *fi2* را دارد، ضرب آن ها یعنی ضرب نقطه ای در فضای ویژگی یعنی ضرب نقطه ای *fi1* و *fi2* که مشخصا معتبر است.

۳- میدانیم که :

$$e^{k_1(x_1, x_2)} = \lim_{i \to \infty} 1 + \frac{k_1(x_1, x_2)}{1!} + \frac{(k_1(x_1, x_2))^2}{2!} + \cdots + \frac{(k_1(x_1, x_2))^i}{i!}$$

از آن جایی که میدانیم اگر *f* یک تابع چند جمله ای با ضرایب مثبت باشد، *f(k(x,z))* نیز معتبر است زیرا هر چندجمله ای ضرب یک سری هسته با ضرایب مثبت است(قسمت ۱ و ۲) لذا *lim* مربوطه هم این ویژگی را دارد و معتبر است. پس هسته مربوطه نیز معتبر است.

۴- میدانیم:

$$(1 - x_1^T x_2)^{-1} = (1 - k_7(x_1, x_2))^{-1} = \frac{1}{1 - k_7(x_1, x_2)}$$

از قانون بسط حسابی میدانیم:

$$\frac{1}{1 - k_7(x_1, x_2)} = \sum_{i=0}^{\infty} k_7^i(x_1, x_2)$$

و از قسمت ۱ میدانیم که جمع کرنل های معتبر با ضریب مثبت نیز معتبر است پس *k6* نیز معتبر است.

سوال سوم – به دلیل ویژگی افزایشی و تقارن، داریم:

$$g(x+y, x+y) = g(x, x+y) + g(y, x+y)$$
$$= g(x+y, x) + g(x+y, y) = g(x,x) + 2g(x,y) + g(y,y)$$

هم چنین داریم:

$$g(x-y, x-y) = g(x, x-y) - g(y, x-y)$$
$$= g(x-y, x) - g(x-y, y) = g(x,x) - 2g(x,y) + g(y,y)$$

لذا با جایگذاری آن ها در رابطه اصلی داریم:

$$h(x,y) = \frac{1}{4}\big(g(x+y, x+y) - g(x-y, x-y)\big)$$
$$= \frac{1}{4}[g(x,x) + 2g(x,y) + g(y,y) - g(x,x) + 2g(x,y) - g(y,y)]$$
$$= \frac{1}{4}[4g(x,y)] = g(x,y)$$

و چون h=g شد، یعنی h نیز ویژگی های گفته شده را دارد.و چون این ویژگی ها، ویژگی های دست بالا برای معتبر بودن یک هسته است، لذا h نیز معتبر است.

سوال چهارم – برای اثبات اینکه K یک هسته معتبر است، فرض میکنیم $\varphi(x)$ یک بردار متناهی است که هر درایه آن متناظر با یک عضو از مجموعه مرجع است. هر درایه ۱ است اگر و تنها اگر عضو متناظر آن در مجموعه مرجع x باشد و در غیر این صورت صفر است.حال میدانیم ضرب داخلی $\varphi(x)$ و $\varphi(y)$ در اصل همان تعداد اعضای مجموعه اشتراک x و z است چون اگر هر دو درایه متناظر غیر صفر باشند، ضرب داخلی آن ها هم غیر صفر میشود و اگر حتی یکی از آن ها صفر باشد، ضرب داخلی آن ها صفر میشود پس مفهوم اشتراک را میرسانند.بنابراین $\varphi(x)\ \ dot\ \ \varphi(y) = |x \cap y|$

حال میدانیم که:

$$2^{|x \cap y|} = e^{\ln 2\, | x \cap y}$$

و از طرفی چون مقدار لگاریتم 2 در پایه طبیعی، یک عدد ثابت مثبت است، هسته $\ln 2\, |\, x \cap y$ یک هسته معتبر است بنابراین هسته k هم یک هسته معتبر است.

مرجع: درس موضوعات پیشرفته در یادگیری ماشین دانشگاه توبینگن

سوال پنجم – همه چیز در نوت بوک آورده شده است.