Prob4

A-session

yes, K_3 is a kernel function.

- K_1/K_2 satisfies the symmetry requirement.
- Since K_1/K_2 is a kernel function, its coresponding matrix is semi-positive def.

  $\forall$ none-zero vector x , $x^T A_1 x \geq 0 \quad x^T A_2 x \geq 0$ *Matrix* A_1/A_2 *is the coresponding matrix of* K_1

  Add them up, we have $x^T(A_1 + A_2)x \geq 0$, indicating K_3 is a kernel.

Hence, $K_3 = K_1 + K_2$, satisfies two requirement of being a valid kernel function.

B-session

No, it's not li-separable for K_3

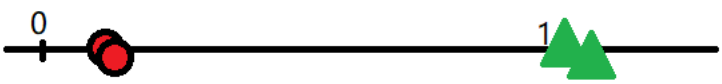
If it's li-separable for K_3, with the given statement, i.e. it's li-separable for K_1, then following the above reasoning, for $K_2 = K_3 - K_1$, it is li-separable, which is false with known information.

Prob5

Gaussian Kernel brings Zero training error. (λ is set to be 1, which can be other value as well.)

$K(a, b) = e^{-||b-a||}$

$x \| (x_1, x_2)$	$K(x_1, x_2)$	$(K(x), y)$
(-1, -1)	e^0	(1,-1)
(-1, +1)	e^{-2}	(e^{-2}, 1)
(+1, -1)	e^{-2}	(e^{-2}, 1)
(+1, +1)	e^0	(1,-1)



Prob 7

How many updates will the algorithm perform, **at the least**?

Only one update(the first time) for the **BEST** condition. (declare $sign(0) = -1$)

We can build a scenario(Best one brings the least updates),

- in which **all** positve samples are $(x, y) \mid (x, y) = (x, 1)\ x \in R^{d^+}$; **all** negative samples are $(x, y) \mid (x, y) = (x, -1)\ x \in R^{d^-}$
- update procedure:
  - round1:
    - $w = 0^d$, sample is $(1^d, 1)$ --> $\hat{y} = sign(w^T x) = -1 \neq y_1$ --> **update x1**
    - $w := w + x$ then w is updated to be 1^d
  - Right now, the w is able to classify samples we **intentionally** built, since:
    - $\forall (x, y) \mid (x, y) = (x, 1)\ x \in R^{d^+}$, $\hat{y} = sign(w^T x) = y$ (y is the label)
    - $\forall (x, y) \mid (x, y) = (x, -1)\ x \in R^{d^-}$, $\hat{y} = sign(w^T x) = y$ (y is the label)

    Hence, no more update is needed.

Based on the 'best' condition, we assert that **at the least**, 1 update is needed.