It can be shown that for any dataset, we can map the original input features to a higher dimensional feature space such that the positive and negative examples are linearly separable in the mapped space. What are the implications of such feature mapping?

A. Feature engineering is no longer important because feature mapping can create the separation that we need.
B. Feature engineering is still important because without the right features, the separation in the mapped space will not be generalizable.

Answer: B

The kernel trick uses a kernel function instead of explicitly mapping the features then taking the dot product. Among the following statements, which ones are true?

A. Using kernel functions (as opposed to mapping then dot product) is a faster way to compute the dot product in the mapped space.
B. Using kernel functions (as opposed to mapping then dot product) is more sample efficient, i.e., it needs fewer training examples to learn effectively.
C. Using kernel functions (as opposed to mapping then dot product) allows us to consider mappings that are not explicitly possible, e.g., due to infinite dimensions.
D. When we use kernel functions as opposed to mapping then dot product, the optimization problem associated with learning can be more efficiently solved.

Answers: A, C

Given a linear decision boundary defined by w and b, the functional margin of a training data point (x,y) is defined as: $y(w^T x + b)$ whereas its geometric margin is defined as $\frac{y(w^T x + b)}{|w|}$, which normalizes the functional margin by the norm of the weight vector w. Which of the following statements are true:

A. Functional margin could be positive or negative. Positive indicates correct prediction and negative indicates incorrect prediction.
B. Geometric margin could be positive or negative. Positive indicates correct prediction and negative indicates incorrect prediction.
C. Consider classifying two training data point with the same classifier (aka, same w and b), if point A has a larger functional margin than point B, point A must be further away from the decision boundary.
D. Consider classifying two training data point with the same classifier (aka, same w and b), if point A has a larger geometric margin than point B, point A must be further away from the decision boundary.

Answers: A, B

Answers C and D are not correct because both margins are signed. Point B could be closer to the decision boundary but on the wrong side.

As we change the parameter C for soft-margin SVM, what do you expect to happen to the number of support vectors?

A. Larger C value leads to larger number of support vectors.
B. Larger C value leads to smaller number of support vectors.
C. If C=0, all points become support vectors.
D. If C= ∞ , all points become support vectors.

Answers: B, C