1. The fundamental idea is to come up with a decision boundary that can separate the classe s

Not only a decision boundary is needed, but the central idea is to maximize the margin between the decision boundary and the two classes.

Soft margin classification: allow some instances on the 'street'

Hard margin classification: do not allow exceptions

2. Support vectors are the instances that are on the "edge of the street". These determine th e decision boundary.

Also includes those "on the street"

3. The SVMs are sensitive to the scaling. Thus, when inputs of one feature is in general great er than that of another feature, the smaller inputs will be considered less for the decision bou ndary.

4. No. It is either a 'yes' or a 'no.'

You can use the distance from the decision boundary to calculate the confidence score. This is done by seeing `probability=True` when creating SVM.

5. The dual problem is faster to solve than the primal one when the number of training instan ces is smaller than the number of features (p. 169). Hence, the primal form is appropriate in t his question.

Additionally, primal form cannot be used in kernelized SVM settings.

6. RBF kernel - Increase C, Increase gamma.

C - Allow less numbers of margin violations (may generalize worse)

gamma - smaller width of the bell-shaped curve - Allow higher degrees of polynomial (may c

ause overfitting)

7. f != 0, but a np-dimensional vector

H: H with m columns of 0 to the right and bottom

F: m additional elements = C

B: m additional elements = 0

A: A with m x m identity matrix to the right and its negative right below it.