Step1

```
SVM kernel=Linear C=0.0001 acc=94.0828402366864 n=72
SVM kernel=Linear C=0.05 acc=92.3076923076923 n=44
SVM kernel=Linear C=0.1 acc=92.89940828402366 n=43
SVM kernel=Linear C=2 acc=95.26627218934911 n=36
SVM kernel=Linear C=5 acc=93.49112426035504 n=34
```

With small values of C, our model will be more generalizable but as C gets smaller values the risk of underfitting increases. On the other hand, as C gets larger values our model can memorize the training so, we can face the risk of overfitting. In our model, the best choice for the C value can be 2.

Support Vectors are the points within the hyperplane margin. For smaller values of C, the margin separating the hyperplane will be large. Since the margin is large, we will have more data points within its range. As C gets greater values, the hyperplane margin will be smaller which causes a smaller number of support vectors.

Step2

```
SVM kernel=Linear C=4 acc=94.67455621301775 n=36
SVM kernel=Polynomial C=4 acc=92.3076923076923 n=26
SVM kernel=RBF C=4 acc=76.92307692307693 n=400
SVM kernel=Sigmoid C=4 acc=76.92307692307693 n=346
```

The best kernel choice for our data is Linear Kernel as its accuracy is the highest one. RBF and Sigmoid Kernels gave us the worst accuracies. RBF generally can be the best choice but since our data is linearly separable, it did not fit to our model.