With supervised learning, the performance of many supervised learning algorithms will be pretty similar,

and **what matters less** often will be whether you use learning algorithm a or learning algorithm b,

but **what matters more** will often be things like the amount of data you create these algorithms on, as well as your skill in applying these algorithms.

[And compared to both logistic regression and neural networks,](https://www.coursera.org/learn/machine-learning/lecture/sHfVT/optimization-objective) [the Support Vector Machine, or SVM sometimes gives a cleaner, and](https://www.coursera.org/learn/machine-learning/lecture/sHfVT/optimization-objective) [sometimes more powerful way of learning complex non-linear functions.](https://www.coursera.org/learn/machine-learning/lecture/sHfVT/optimization-objective)

Support Vector Machine, is a supervisory algorithm. Modify a bit logistic regression, get what is essentially the support vector machine.

When y =1 then we're sort of hoping that h(x) will be close to one. Right, we're hoping to correctly classify that example. Conversely, if we have an example where y=0, then what we're hoping for is that the hypothesis will output a value close to zero.

Unlike logistic regression, the support vector machine doesn't output the probability is that what we have is we have this cost function, that we minimize to get the parameter's data, and what a support vector machine does is it just makes a prediction of y being equal to one or zero, directly. So the hypothesis will predict one

If your algorithm suffering from **a high bias problem**, then your **training and cross-validation set error will be high** as your hypothesis just not fitting training set well. Decrease σ2 and increase C

If your algorithm suffering from **a high variance problem**, then your **training set error will be low** as your fitting training set well but **cross-validation set error** will be high and you may be over fitting and regularization may help. Increase σ2 and decrease C.

Large σ2 - f features vary more smoothly - higher bias, lower variance

Small σ2 - f features vary abruptly - low bias, high variance

Large C gives a hypothesis of **low bias high variance** --> overfitting

Small C gives a hypothesis of **high bias low variance** --> underfitting