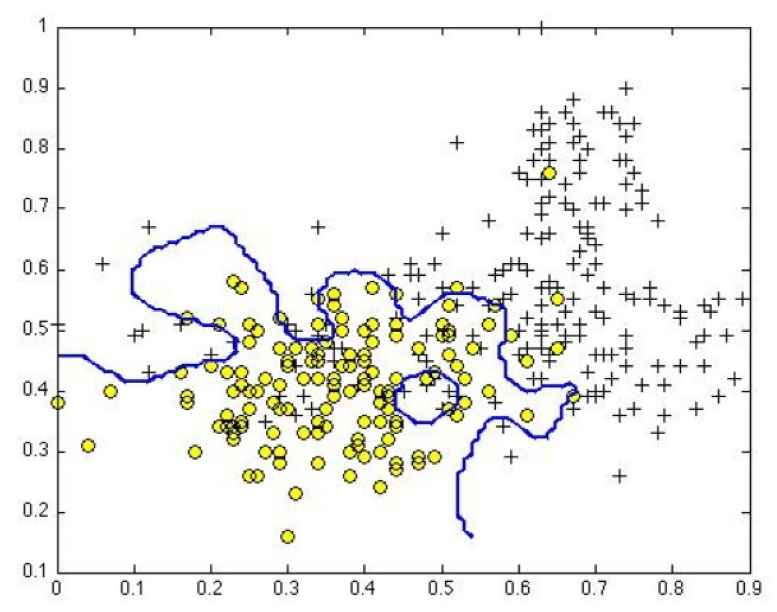
Support Vector Machines

Quiz, 5 questions

1 point

1. Suppose you have trained an SVM classifier with a Gaussian kernel, and it learned the following decision boundary on the training set:



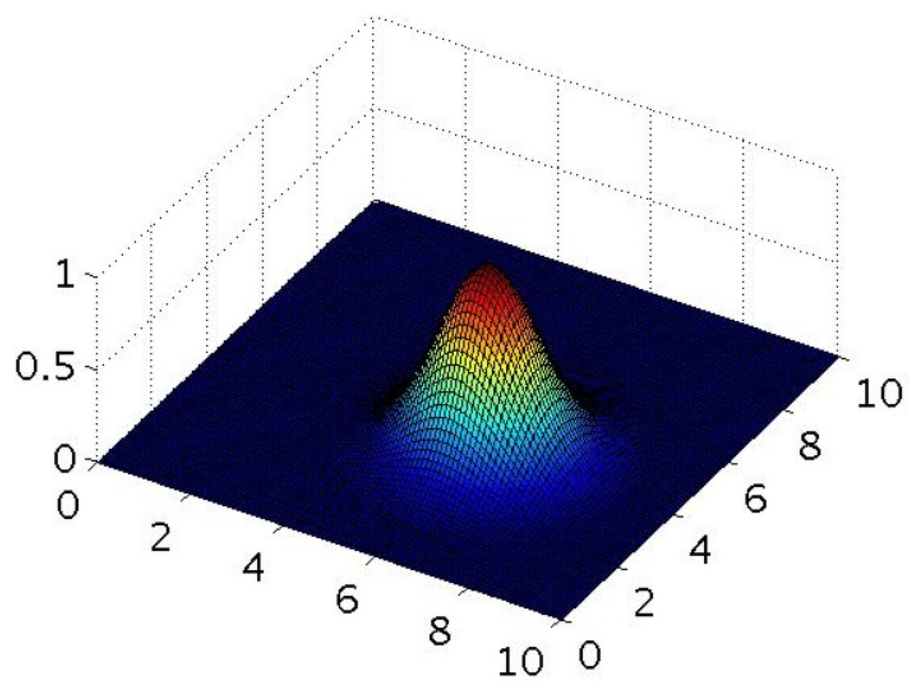
When you measure the SVM's performance on a cross validation set, it does poorly. Should you try increasing or decreasing C? Increasing or decreasing σ^2?

- [ ] It would be reasonable to try **increasing** C. It would also be reasonable to try **increasing** σ^2.
- [x] It would be reasonable to try **decreasing** C. It would also be reasonable to try **increasing** σ^2.
- [ ] It would be reasonable to try **decreasing** C. It would also be reasonable to try **decreasing** σ^2.
- [ ] It would be reasonable to try **increasing** C. It would also be reasonable to try **decreasing** σ^2.

1 point

2. The formula for the Gaussian kernel is given by $\mathrm{similarity}(x, l^{(1)}) = \exp\left(-\frac{||x-l^{(1)}||^2}{2\sigma^2}\right)$.

The figure below shows a plot of $f_1 = \mathrm{similarity}(x, l^{(1)})$ when $\sigma^2 = 1$.



Which of the following is a plot of f_1 when $\sigma^2 = 0.25$?

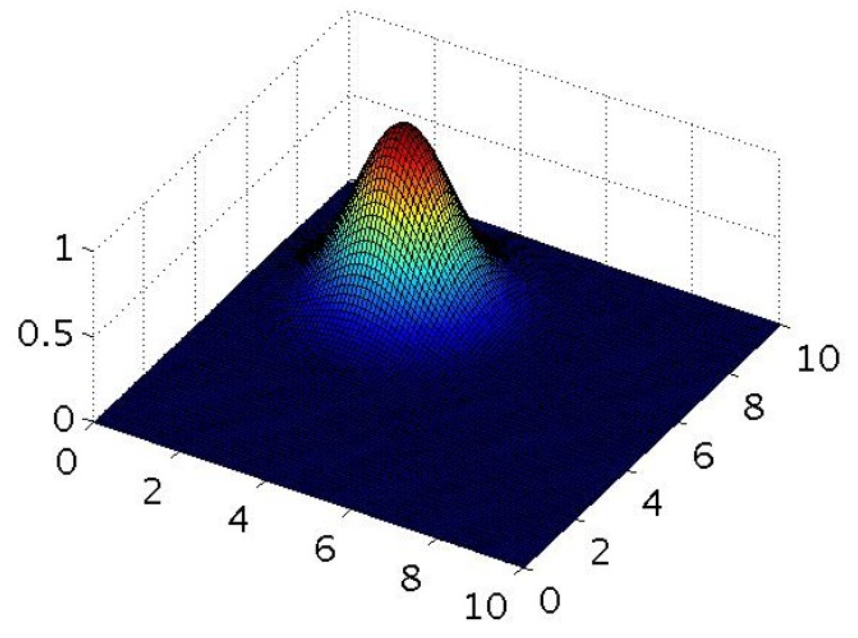
- [ ] 

Figure 4.

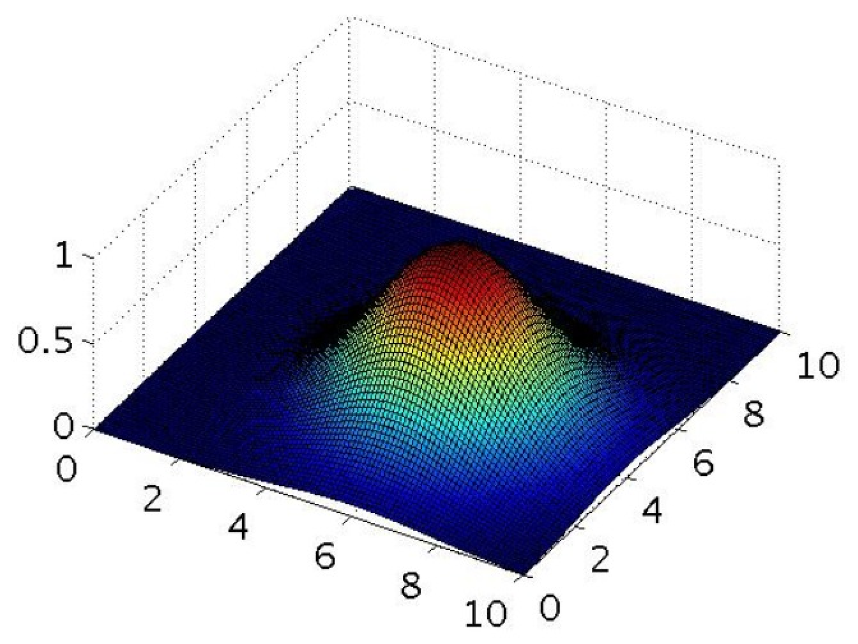
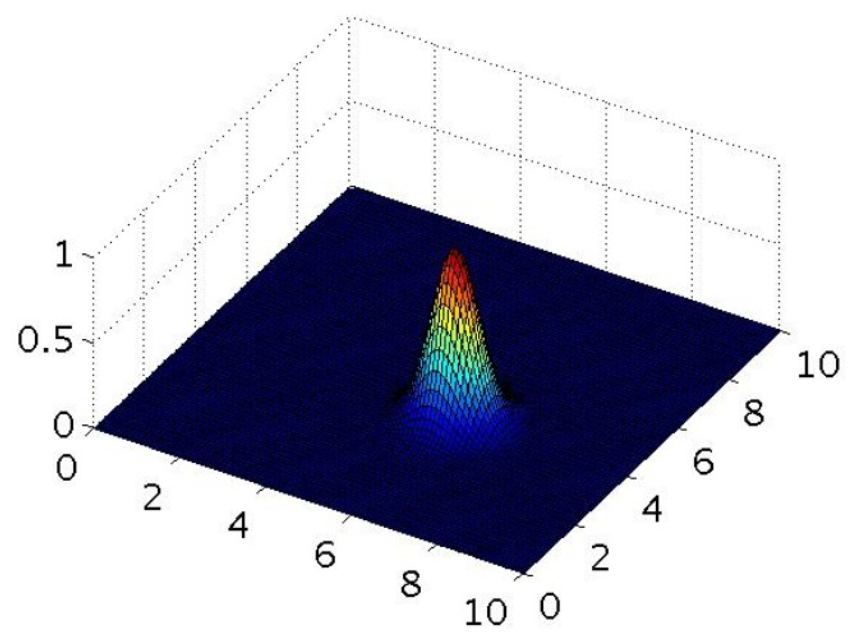
- [ ] 

Figure 2.

- [x] 

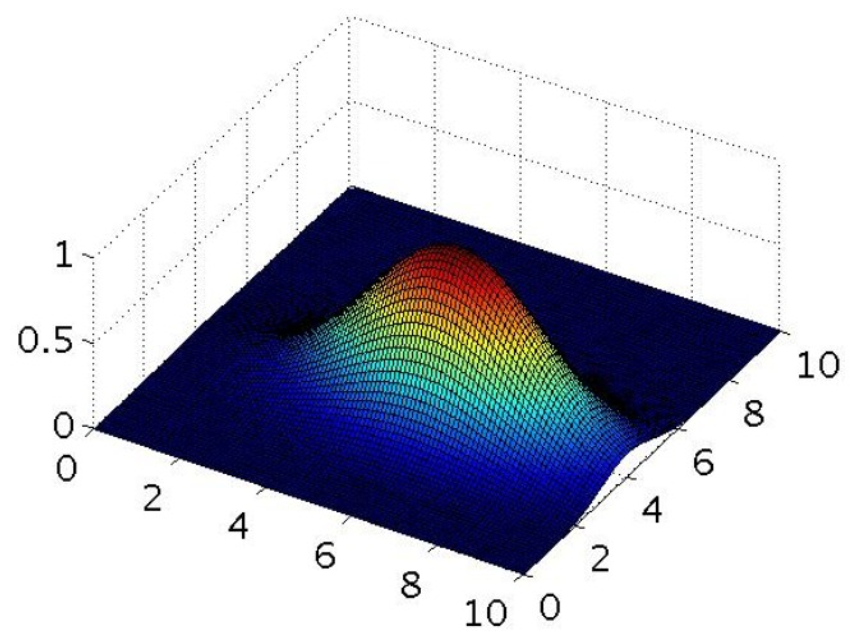
- [ ] 

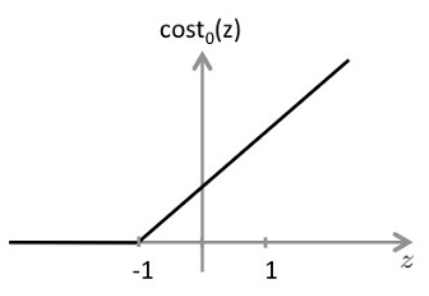
Figure 3.

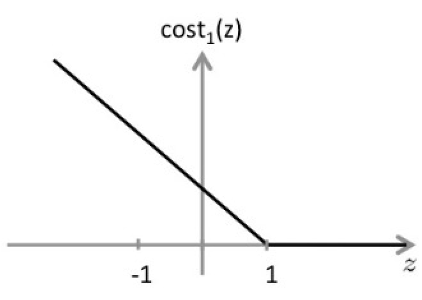
1 point

3. The SVM solves

$\min_\theta\ C\sum_{i=1}^{m} y^{(i)}\mathrm{cost}_1(\theta^T x^{(i)}) + (1-y^{(i)})\mathrm{cost}_0(\theta^T x^{(i)}) + \sum_{j=1}^{n}\theta_j^2$

where the functions $\mathrm{cost}_0(z)$ and $\mathrm{cost}_1(z)$ look like this:





The first term in the objective is:

$C\sum_{i=1}^{m} y^{(i)}\mathrm{cost}_1(\theta^T x^{(i)}) + (1-y^{(i)})\mathrm{cost}_0(\theta^T x^{(i)})$.

This first term will be zero if two of the following four conditions hold true. Which are the two conditions that would guarantee that this term equals zero?

- [ ] For every example with $y^{(i)} = 0$, we have that $\theta^T x^{(i)} \leq 0$.
- [x] For every example with $y^{(i)} = 0$, we have that $\theta^T x^{(i)} \leq -1$.
- [x] For every example with $y^{(i)} = 1$, we have that $\theta^T x^{(i)} \geq 1$.
- [ ] For every example with $y^{(i)} = 1$, we have that $\theta^T x^{(i)} \geq 0$.

1 point

4. Suppose you have a dataset with n = 10 features and m = 5000 examples.

After training your logistic regression classifier with gradient descent, you find that it has underfit the training set and does not achieve the desired performance on the training or cross validation sets.

Which of the following might be promising steps to take? Check all that apply.

- [ ] Reduce the number of examples in the training set.
- [x] Try using a neural network with a large number of hidden units.
- [x] Create / add new polynomial features.
- [ ] Use a different optimization method since using gradient descent to train logistic regression might result in a local minimum.

1 point

5. Which of the following statements are true? Check all that apply.

- [ ] If the data are linearly separable, an SVM using a linear kernel will

  return the same parameters θ regardless of the chosen value of

  C (i.e., the resulting value of θ does not depend on C).
- [ ] Suppose you have 2D input examples (ie, $x^{(i)} \in \mathbb{R}^2$). The decision boundary of the SVM (with the linear kernel) is a straight line.
- [x] The maximum value of the Gaussian kernel (i.e., $sim(x, l^{(1)})$) is 1.
- [ ] If you are training multi-class SVMs with the one-vs-all method, it is

  not possible to use a kernel.

- [x] I understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code

Attila Földvárszky

Use the name on your government issued ID

Submit Quiz