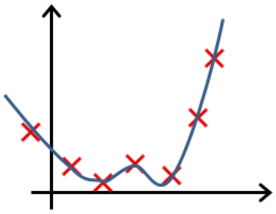
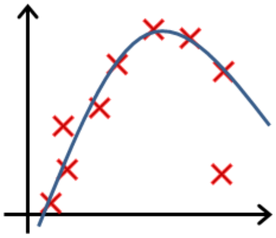
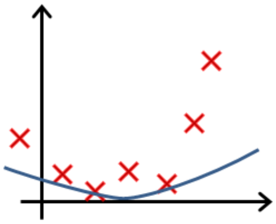
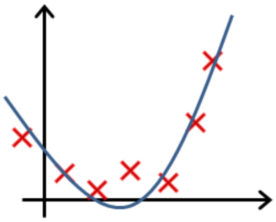
Feedback — VII. Regularization

Help

You submitted this quiz on **Sat 12 Jul 2014 3:21 AM PDT**. You got a score of **5.00** out of **5.00**.

Question 1

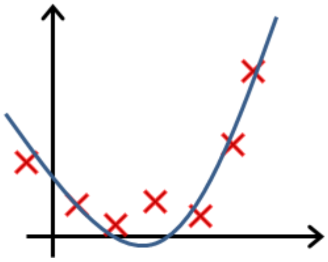
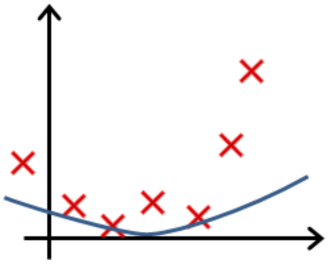
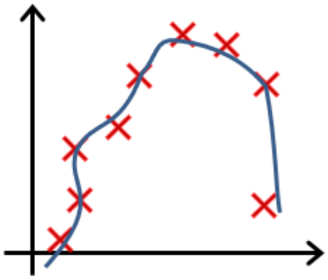
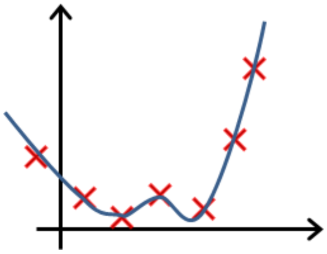
In which one of the following figures do you think the hypothesis has overfit the training set?

Your Answer		Score	Explanation
● 	✓	1.00	The hypothesis follows the data points very closely and is highly complicated, indicating that it is overfitting the training set.
○ 			
○ 			
○ 			
Total		1.00 /	

1.00

Question 2

In which one of the following figures do you think the hypothesis has underfit the training set?

Your Answer		Score	Explanation
○ 			
● 	✓	1.00	The hypothesis does not predict many data points well, and is thus underfitting the training set.
○ 			
○ 			
Total		1.00 / 1.00	

Question 3

You are training a classification model with logistic regression. Which of the following statements are true? Check all that apply.

Your Answer		Score	Explanation
☑ Adding a new feature to the model always results in equal or better performance on the training set.	✓	0.25	By adding a new feature, our model must be more (or just as) expressive, thus allowing it learn more complex hypotheses to fit the training set.
☐ Introducing regularization to the model always results in equal or better performance on the training set.	✓	0.25	If we introduce too much regularization, we can underfit the training set and have worse performance on the training set.
☐ Introducing regularization to the model always results in equal or better performance on examples not in the training set.	✓	0.25	If we introduce too much regularization, we can underfit the training set and this can lead to worse performance even for examples not in the training set.
☐ Adding a new feature to the model always results in equal or better performance on examples not in the training set.	✓	0.25	Adding more features might result in a model that overfits the training set, and thus can lead to worse performs for examples which are not in the training set.
Total		1.00 / 1.00	

Question 4

Suppose you ran logistic regression twice, once with $\lambda = 0$, and once with $\lambda = 1$. One of the times, you got parameters $\theta = \begin{bmatrix} 74.81 \\ 45.05 \end{bmatrix}$, and the other time you got $\theta = \begin{bmatrix} 1.37 \\ 0.51 \end{bmatrix}$. However, you forgot which value of λ corresponds to which value of θ. Which one do you think corresponds to $\lambda = 1$?

Your Answer		Score	Explanation
● $\theta = \begin{bmatrix} 1.37 \\ 0.51 \end{bmatrix}$	✓	1.00	When λ is set to 1, we use regularization to penalize large values of θ. Thus, the parameters, θ, obtained will in general have smaller values.
○ $\theta = \begin{bmatrix} 74.81 \\ 45.05 \end{bmatrix}$			
Total		1.00 / 1.00	

Question 5

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

Your Answer		Score	Explanation
☐ Using a very large value of λ cannot hurt the performance of your hypothesis; the only reason we do not set λ to be too large is to avoid numerical problems.	✓	0.25	Using a very large value of λ can lead to underfitting of the training set.
☐ Because logistic regression outputs values $0 \leq h_\theta(x) \leq 1$, it's range of output values can only be "shrunk" slightly by regularization anyway, so regularization is generally not helpful for it.	✓	0.25	Regularization affects the parameters θ and is also helpful for logistic regression.

☑ Consider a classification problem. Adding regularization may cause your classifier to incorrectly classify some training examples (which it had correctly classified when not using regularization, i.e. when $\lambda = 0$).	✓	0.25	Regularization penalizes complex models (with large values of θ).They can lead to a simpler models, which misclassifies more training examples.
☐ Because regularization causes $J(\theta)$ to no longer be convex, gradient descent may not always converge to the global minimum (when $\lambda > 0$, and when using an appropriate learning rate α).	✓	0.25	Regularized logistic regression and regularized linear regression are both convex, and thus gradient descent will still converge to the global minimum.
Total		1.00 / 1.00	