Question 2a Recall the optimal value of θ should minimize our loss function. One way we've approached solving for θ is by taking the derivative of our loss function with respect to θ, like we did in HW5.

Write/derive the expressions for following values and write them with LaTeX in the space below.

- $R(\mathbf{x}, \mathbf{y}, \theta_1, \theta_2)$: our loss function, the empirical risk/mean squared error
- $\frac{\partial R}{\partial \theta_1}$: the partial derivative of R with respect to θ_1
- $\frac{\partial R}{\partial \theta_2}$: the partial derivative of R with respect to θ_2

Recall that $R(\mathbf{x}, \mathbf{y}, \theta_1, \theta_2) = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2$

- $R(\mathbf{x}, \mathbf{y}, \theta_1, \theta_2) = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2$
- $\hat{\mathbf{y}} = \theta_1 x + sin(\theta_2 x)$

$\frac{\partial R}{\partial \theta_1} = \frac{1}{n} \sum_{i=1}^{n} (\theta_1 \mathbf{x}_i + sin(\theta_2 \mathbf{x}_i) - \mathbf{y}_i) * 2\mathbf{x}_i = \frac{1}{n} \sum_{i=1}^{n} ((\hat{\mathbf{y}} - \mathbf{y}_i) * 2\mathbf{x}_i)$

$\frac{\partial R}{\partial \theta_2} = \frac{1}{n} \sum_{i=1}^{n} (sin(\theta_2 \mathbf{x}_i) + \theta_1 \mathbf{x}_i - \mathbf{y}_i) * cos(\theta_2 \mathbf{x}_i) * 2\mathbf{x}_i = \frac{1}{n} \sum_{i=1}^{n} ((\hat{\mathbf{y}} - \mathbf{y}_i) * 2\mathbf{x}_i cos(\theta_2 \mathbf{x}_i))$

In 1-2 sentences, describe what you notice about the path that theta takes with a static learning rate vs. a decaying learning rate. In your answer, refer to either pair of plots above (the 3d plot or the contour plot).

Static learning rate has much more turning point(jumping left and right). Each successive gradient jumps in all directions.
In contrast, decaying learning rate is more stable. The gradient generally descends in one direction.

0.0.1 Question 4b

Is this model reasonable? Why or why not?

This model is not reasonable. The team has a high probability(>=50%) to win no matter how many points they got, which means probability is not related to score. This will cause problems for our analysis.

0.0.2 Question 4c

Try playing around with other theta values. You should observe that the models are all pretty bad, no matter what θ you pick. Explain why below.

Because this is binary classification. Our responses are either 0(lost) or 1(win), they have no linear relationship.

0.0.3 Question 5b

Using the plot above, try adjusting θ_2 (only). Describe how changing θ_2 affects the prediction curve. Provide your description in the cell below.

As θ_2 increases, the symmetry axis of the image will move to the left, which mean all the win probability will increase, up to 1.

On the contrary, as θ_2 decreases, the symmetry axis will move to the right, which mean all the win probability will decrease, down to 0.

0.0.4 Question 7c

Look at the coefficients in `theta_19_hat` and identify which of the parameters have the biggest effect on the prediction. For this, you might find `useful_numeric_fields.columns` useful. Which attributes have the biggest positive effect on a team's success? The biggest negative effects? Do the results surprise you?

In [68]:
```
theta_19_hat,useful_numeric_fields.columns
```

Out[68]:
```
(array([ 2.1239e+00, -4.5115e-01, -2.1898e+01,  9.2304e-01, -3.5553e-03,
         2.7252e+00,  8.8270e-01, -7.1648e-02,  2.1081e+00,  3.0481e-01,
         3.2550e-01,  4.8715e-02,  1.9556e-02,  3.8726e-01,  6.9643e-02,
        -3.1119e-01, -5.5789e-02, -7.5458e-01,  5.1914e+00]),
 Index(['FGM', 'FGA', 'FG_PCT', 'FG3M', 'FG3A', 'FG3_PCT', 'FTM', 'FTA',
        'FT_PCT', 'OREB', 'DREB', 'REB', 'AST', 'STL', 'BLK', 'TOV', 'PF',
        'PTS', 'BIAS'],
       dtype='object'))
```

FG3_PCT, FGM, FT_PCT have the biggest effect and excluding the (intercept)bias term and FG_PCT, PTS, FGA have the biggest negative effects. I'm not surprised by this because it seems reasonable. 3 point field goal percentage, field goals made, free throw percentage can help to improve the score and thus the predicted value. Field goal percentage and attempted have negative effects because make/attempted+make = PCT in the case of the same make, the more attempts, the more mistakes and cause the team to be unable to score.

To double-check your work, the cell below will rerun all of the autograder tests.

In [65]:
```
grader.check_all()
```

Out[65]:
```
q1:

    All tests passed!

q2b:

    All tests passed!

q3a:

    All tests passed!
```

```
q3b:

    All tests passed!

q4a:

    All tests passed!

q5a:

    All tests passed!

q5c:

    All tests passed!

q6a:

    All tests passed!

q6b:

    All tests passed!

q6c:

    All tests passed!

q6d:

    All tests passed!

q7a:

    All tests passed!

q7b:

    All tests passed!
```

0.1 Submission

Make sure you have run all cells in your notebook in order before running the cell below, so that all images/graphs appear in the output. The cell below will generate a zip file for you to submit. **Please save before exporting!**

In [66]:
```
# Save your notebook first, then run this cell to export your submission.
grader.export("hw7.ipynb")
```

```
<IPython.core.display.HTML object>
```