**Question 2a** Recall the optimal value of $\theta$ should minimize our loss function. One way we've approached solving for $\theta$ is by taking the derivative of our loss function with respect to $\theta$, 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 $\theta_1$
- $\frac{\partial R}{\partial \theta_2}$: the partial derivative of $R$ with respect to $\theta_2$

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

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

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

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

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).

In case of the static learning rate, since alpha stays constant even when theta approaches the minimum, we see that theta can potentially oscillate around the minima for a large number of iterations before reaching the minima. This is visible in the 3D plot where theta quickly descends due to the static learning rate but has trouble reaching the minima.

As opposed to that, the decaying learning rate does not allow for too much oscillation. Theta takes a more linear path to the minima which could be beneficial but due to the reduced learning rate, it takes longer for theta to reach the minima.

### 0.0.1 Question 4b

Is this model reasonable? Why or why not?

No it is not reasonable. This linear model is not taking into account the fact that only 2 values are possible and is very suseptible to outliers. Also the model has a narrow range of 0.5 to 0.8 in probability. This model is lacking an intercept and is making the assumption that probability of receiving 0 points is 0.5 which does not sound right.

### 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 $\theta$ you pick. Explain why below.

Once again, this error is caused due to the lack of an intercept. For any value of theta, the model predicts the probability of getting 0 points at 0.5 which is not correct. This is because the model is forced to pass through the origin, even if it may not yield the best results.

### 0.0.3 Question 5b

Using the plot above, try adjusting $\theta_2$ (only). Describe how changing $\theta_2$ affects the prediction curve. Provide your description in the cell below.

As the value od $\theta_2$ increases, the curve shifts to the left. But the plot shows that even with $\theta_1$ constant the curves are not parallel. The predicted value for each point reduces as $\theta_2$ increases. For instance, different values of $\theta_2$ have been plotted above and if we consider point x=100 the corresponding points on the blue, orange and pink curves are 0.7,0.5 and 0.3 respectively.

### 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 [165]:
```
use_num = useful_numeric_fields.columns.to_frame()
use_num["weights"] = theta_19_hat
use_num.sort_values("weights")
```

Out[165]:

| | 0 | weights |
|---|---|---|
| FG_PCT | FG_PCT | -21.898362 |
| PTS | PTS | -0.754584 |
| FGA | FGA | -0.451155 |
| TOV | TOV | -0.311187 |
| FTA | FTA | -0.071648 |
| PF | PF | -0.055789 |
| FG3A | FG3A | -0.003555 |
| AST | AST | 0.019556 |
| REB | REB | 0.048715 |
| BLK | BLK | 0.069643 |
| OREB | OREB | 0.304814 |
| DREB | DREB | 0.325504 |
| STL | STL | 0.387256 |
| FTM | FTM | 0.882696 |
| FG3M | FG3M | 0.923039 |
| FT_PCT | FT_PCT | 2.108145 |
| FGM | FGM | 2.123860 |
| FG3_PCT | FG3_PCT | 2.725184 |
| BIAS | BIAS | 5.191449 |

The column with the most negative effect is FG_PCT which is field goal percentage. The column with most positive effect is the BIAS column which makes sense since the bias term aims to balance the effect of other parameters.

---

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

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

## 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 [129]:
```
# Save your notebook first, then run this cell to export your submission.
grader.export("hw7.ipynb")
```

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