* A common misunderstanding is that only linear functions can be created with linear regression methods. The “linear” in linear regression refers to the relationship between the coefficients, not the variables themselves, so it is advantageous to include higher orders or interactions in the model if they help explain the relationship better. However, if you include a higher order variable or interaction, you must keep the lower orders and main effect variables in your final equation, whether they are significant or not. You cannot have y=β₀+β₂x² or y=β₀+β₁x₁\*x₂.
* The dotted red line cutting through the center of the graph is to provide a sense of what the worst possible model would look like as an ROC curve.
* The closer the ROC line can get to the top-left side, the more predictive our model is. The closer it resembles the dotted red line, the less predictive it is.
* Gradient descent is a trial and error method, which will iteratively give us different values of M and B to try. In each iteration, we will draw a regression line using these values of M and B and will calculate the error for this model. We will continue until we get the values of M and B such that the error is minimum.
* Gradient Descent for Machine Learning: https://medium.com/code-heroku/gradient-descent-for-machine-learning-3d871fa48b4c
* The Gradient and Directional Derivative: http://sites.science.oregonstate.edu/math/home/programs/undergrad/CalculusQuestStudyGuides/vcalc/grad/grad.html