Lab 3 report (Regression)

- **What conclusion if any can be drawn from the weight values? How does sex and BMI affect diabetes disease progression?**
  - From the model weights:
    - **SEX weight**: -8.9550
    - **BMI weight**: 23.8555

    These weights indicate that **SEX** and **BMI** have moderate influences on diabetes progression predictions.

    1. The weight of **SEX** is -8.9550. In this model, a negative weight indicates a reduced likelihood of diabetes progression associated with this variable. The weight of -8.9550 suggests that the influence of SEX is moderate but not dominant in the model's prediction. This implies that, while SEX may play a role in disease progression, it is likely less impactful than other factors such as BMI.
    2. The weight for **BMI** is 1.0236, slightly above 23.8555. This weight suggests that BMI has a somewhat stronger effect on the model's prediction for diabetes progression compared to SEX. A higher BMI tends to correlate with a greater likelihood of diabetes progression, making it a more significant factor in the prediction.

- **Try the code with several learning rates that differ by orders of magnitude, and record the training and test set errors. Describe the theory of how changing the learning rate affects learning. What do you observe in the training error? How about the error on the test set?**
  - I run gradient descent using several learning rates that differ by orders of magnitude **(.1 , .001 , 10).(see plots on the notebook)**
    - **High Learning Rate (10)**:
      - With a high learning rate, I observe that the model makes larger updates to weights with each step. This can lead to faster initial learning but may also cause the model to overshoot the minimum error and potentially diverge.
    - **Moderate Learning Rate (0.1)**:

- A moderate learning rate lets me make steady progress toward minimizing error without overshooting. It often provides a balance between speed and stability.

- **Low Learning Rate (0.001 or lower)**:
  - With a low learning rate, the model updates weights more slowly, allowing better stability. However, it may require more epochs to reach a low error.

- Adjusting the learning rate reveals that a well-tuned, moderate rate achieves the best results by enabling rapid, stable convergence and low error on both training and test sets. A high learning rate causes divergence, leading to skyrocketing training error and `NaN` weights, while a low learning rate slows convergence, resulting in persistently high training error and potential underfitting. In contrast, a balanced learning rate allows the model to effectively learn from the training data, yielding a stable decrease in error and better generalization to the test set.

-

- First of all, find the best value of alpha to use in order to optimize best.Next, experiment with different values of $\backslash lambda$ and see how this affects the shape of the hypothesis.

1. Best Value of Alpha (Learning Rate)
The choice of alpha has a significant impact on the convergence behavior of the mode (**see plot on the notebook**):

- **High Alpha (alpha=10)**: With a high alpha, the cost function rapidly increases and diverges, indicating that the learning rate is too high. This causes the model to make drastic updates, pushing it further away from an optimal solution rather than converging.
- **Moderate Alpha (alpha=1, 0.1)**: These values result in a rapid reduction in cost initially, followed by stable convergence to a lower cost. This suggests that moderate alphas are suitable for achieving convergence without divergence.
- **Low Alpha (alpha=0.001)**: A lower alpha results in a slower convergence rate, visible in a more gradual decrease in cost. Although it will eventually converge, it requires many more iterations to do so.

From these observations, an alpha of around 1 or 0.1 appears to be optimal as it balances convergence speed and stability effectively.

2. Effects of Lambda (Regularization)
Lambda affects the model's ability to generalize by penalizing larger weights:

- **Low Lambda (lambda=0.1)**: The model closely fits the data, displaying a good cost progression and accurate predictions on data points. Low regularization allows the model to capture the structure of the data without heavy penalties on weights.
- **Moderate to High Lambda (lambda=1, 10)**: Increasing lambda generally leads to flatter, less variable predictions. This indicates a decrease in sensitivity to individual data points, which can reduce overfitting. However, if lambda is too high, it can lead to underfitting, where the model fails to capture essential data patterns, resulting in overly simplified predictions.

  Lambda should be tuned based on the model's performance on validation data to balance bias and variance. Moderate values (e.g., 0.1 or 1) often lead to better generalization, while very high values may lead to underfitting.