1. The intercept, TV, and radio all have nearly non-existent p-values (<10^-4), indicating a strong correlation between these three predictors and the response (sales), and a high likelihood that the null hypothesis can be refuted for these three. Conversely, newspaper has a significantly higher and notable p-value of 0.8599; likely denoting how there is little to no correlation between this predictor and the response, thereby affirming the null hypothesis.
2. The KNN classifier identifies the K points in the training data closest to any given test observation, using them to estimate the probability that said observation can be classified as some class “a”, selecting the most likely class for that observation. On the other hand, KNN regression is quite similar to its classifier cousin, but rather than using the K nearest points to estimate probability, it instead uses them to directly estimate the response for a given test input using the average of the responses corresponding to each of the K nearest training inputs.
3. GPA, IQ, Gender (1=Female, 0=Male), Interaction between GPA/IQ, Interaction between GPA/Gender; Salary=50+20X1+0.07X2+35X3+0.01X1X2-10X1X3
   1. The fact that for a fixed value of IQ and GPA, males earn more on average provided with a high enough GPA is correct, as and remain fixed, but the female will gain thousand dollars in this case, while also losing thousand dollars. The only differences in the salaries of a male and a female with the same IQ and GPA are those two terms, so if the amount lost by the female () is greater than the amount they gain (35), or when the GPA of these two people is greater than 3.5, the male will earn more than the female.
   2. The salary of a female with IQ 110 and GPA 4.0 andwould be:
   3. The coefficient for the GPA/IQ interaction term being small does not indicate that there is little evidence for interaction. What determines the presence of synergy /interaction are the p-values and t-statistics of the predictors in the fitted model, and the coefficient of this interaction term being small only indicates that the synergy effect between these two predictors is of little impact, not that there exists little evidence for their interaction.
4. Linear regression VS Cubic regression
   1. If the true relationship between X and Y is linear, then the training RSS for cubic regression will be lower, as there exist more variables in this model, allowing it to more accurately produce outputs for the training data.
   2. However, the test RSS is likely lower for linear regression in this case, as the linear regression will more accurately predict the outputs for data with a linear relationship in practice, especially when compared to cubic regression.
   3. If the true relationship between X and Y is not linear, one would still expect the training RSS of cubic regression to be lower, for the same reason as in a).
   4. If we were to consider test RSS for both of these, though, there would not be enough information to tell. This is because all we know about the true nature of the relationship between X and Y is that it is non-linear. If it was cubic
   5. or of a higher degree, then cubic regression would likely be superior, but if it was quadratic or of a lower degree, or even any other non-linear function type (logarithmic, exponential, etc), it would be impossible to definitively say.
5. Linear regression without an intercept has the formula: ,where . We can rewrite the above formula as:  
   Therefore, can be written in the form where .
6. 3.4 is as follows:

In the case of simple linear regression, there are only the two coefficients of and to worry about, as there is only one predictor x being regressed onto the response. If we rearrange the second equation, we obtain , and since the two coefficients will always be constant, inputting with these two coefficients will always result in the response of. Therefore, the least squares line, having equation , will always pass through the point ( ,).