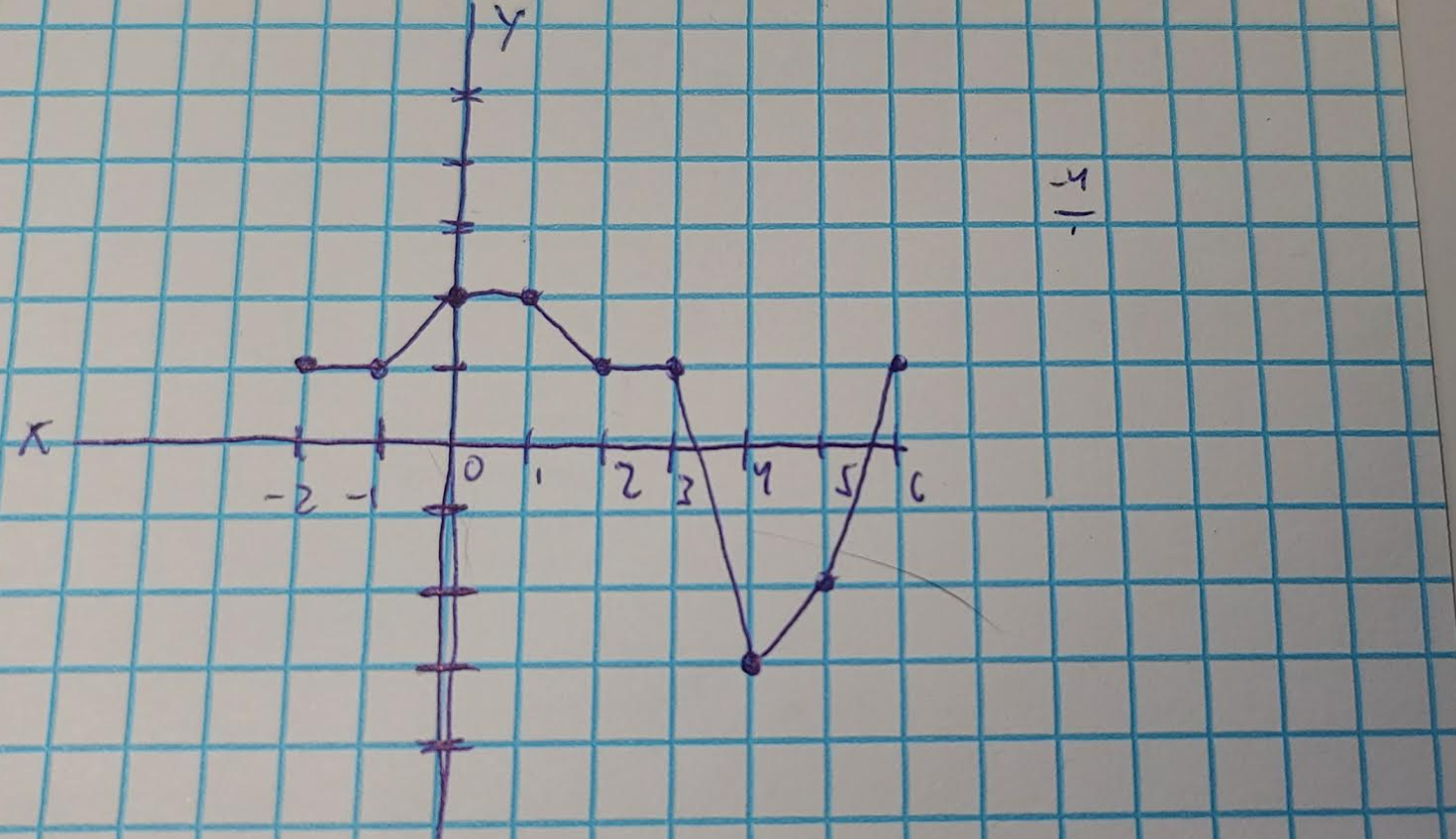
**4. Suppose we fit a curve with basis functions b1(X) = I(0 ≤ X ≤ 2) − (X −1)I(1 ≤ X ≤ 2), b2(X)=(X −3)I(3 ≤ X ≤ 4) +I(4 < X ≤ 5). We fit the linear regression model Y = β0 + β1b1(X) + β2b2(X) + ϵ, and obtain coefficient estimates βˆ0 = 1, βˆ1 = 1, βˆ2 = 3. Sketch the estimated curve between X = −2 and X = 6. Note the intercepts, slopes, and other relevant information.**

****

|  |  |
| --- | --- |
| **X** | **1 + b\_1(x) - 2(b\_2(x)) + e** |
| -2 | 1 |
| -1 | 1 |
| 0 | 2 |
| 1 | 2 |
| 2 | 1 |
| 3 | 1 |
| 4 | -3 |
| 5 | -1 |
| 6 | 1 |

This curve intercepts at the Y axis at x = 2 and intercepts the x axis at a value between x = 3 and x = 4, then again at a value between x = 5 and x = 6. There are several areas where it appears we have a slope of 0. These intervals are [-2, -1], [0, 1], and [2, 3]. At x = 4 we dip hard into the negative values. The slope between x = 3 and x = 4 is -4. It appears after x = 4 we do gradually rise back into positive values. Without more granular x values my estimation on the shape of this curve is a slightly offset and exaggerated cosine wave. At least between x = -2 and x = 3. After that it does lose the characteristic wave appearance cosine waves have, however this curve could be more like a polynomial with an odd degree. Our max value is 2 on the interval of [0, 1] and our min value is at x = 4 which is -3.

**5. Consider two curves, ˆg1 and ˆg2, defined by**

**Text, letter

Description automatically generated**

**where g(m) represents the mth derivative of g.**

**(a) As λ → ∞, will ˆg1 or ˆg2 have the smaller training RSS?**

As lambda approaches infinity we should see g3(x) approach 0. This would indicate that g1 is quadradic of some form.

Similarly with g4(x) we see that lambda approaching infinity should have it approach 0 as well. This then indicates again that g2 is a cubic function of some form.

Since cubic functions are much more flexible in comparison to a quadratic function we should see a smaller training RSS on the cubic function. So g2 should have smaller training RSS

**(b) As λ → ∞, will ˆg1 or ˆg2 have the smaller test RSS?**

This is more difficult to answer. Since we do not know how well either of these fit, we can’t say if one will have smaller test RSS than the other. It is possible that g2 does not overfit the data and thus has a smaller RSS. Or perhaps g2 does overfit the data making g1 a better fit.

**(c) For λ = 0, will ˆg1 or ˆg2 have the smaller training and test RSS?**

At lambda = 0 we can say for certain that the training and test RSS will be the same. This is because the only terms that differ for g1 and g2 zero out at lambda = 0. Essentially at 0 both g1 and g2 are the same.