P1.

I would choose to use k = 10 as it yields the smallest error rate based on the plot. Need to check the assumption that the amount positive class is close to the amount of negative class in the dataset.

First thing is that the approach fails to normalize/rescale the data before fitting the model. Therefore, one step to normalize the data should be inserted right before step3. Second thing is that instead of validation error, the approach should use AUC as the metric to evaluate the model prediction performance. As AUC curve tells the model's marginal prediction performance within each class, which generalizes the performance estimate to the entire population. For example, given a huge dataset, 90% of it are positive classes and 10% are negative classes. By simply saying that the dataset is all positive would be enough for a prediction model to give a satisfactory accuracy score of around 90%, or validation error of around 10%. Therefore, in step4, the approach should use AUC curve rather than validation error to estimate model prediction. Lastly and most importantly, step 2 should be relocated right before step 1. And the predictor screening processing should be done only in the training dataset after splitting the data. Otherwise there the validation prediction would be biased.

P2.

Without scaling:

$$\underset{b}{\operatorname{argmin}}(y - xb)^2 + \lambda b^2$$

Taking the derivative of it with respect to b and set it equals to 0 gives:

$$-2yx + 2x^2b + 2\lambda b = 0$$
$$b = \frac{yx}{x^2 + \lambda}$$

With scaling:

$$\underset{b}{\operatorname{argmin}}(y^* - x^*b)^2 + \lambda^* b^2$$
$$= \underset{b}{\operatorname{argmin}}(ay - axb)^2 + \lambda^* b^2$$

Again, taking the derivative of it with respect to b and set it equals to 0 gives:

$$-2a^2yx + 2a^2x^2b + 2\lambda^* b = 0$$

$$b = \frac{a^2 yx}{a^2 x^2 + \lambda^*}$$

Set it equal to $\frac{yx}{x^2+\lambda}$ we have:

$$\frac{a^2 yx}{a^2 x^2 + \lambda^*} = \frac{yx}{x^2 + \lambda}$$
$$\to a^2 yx^3 + a^2 yx\lambda = a^2 yx^3 + yx\lambda^*$$
$$\to \lambda^* = a^2 \lambda$$

As β stays unchanged, an upward scaling ($\alpha > 1$) would increase the penalty, whereas a downward scaling ($\alpha < 1$) would decrease the penalty.