1.

(a) Overfitting occurs when a machine learning model overcompensates for the variation in data which can be the result of noise, and often results in negative performance as the model will have more extreme fluctuations across its predictions. Two ways to alleviate are to use a loss function and use regularization which both quantify how well the model fits data, and determine which models better fit the data than others.

(b) The magnitude of K impact's the number of nearest neighbours of a data point when a model is classifying a large section of data. A greater value results in more smoother boundaries but can underfit, whereas a smaller value will have significantly sharper decision boundaries, but can overfit. The best way to determine the optimal K value is to simply rerun and train the model using different K values, and evaluating the overall performance of these variations based on classification errors.

(c) An application for regression could be used for looking at multiple features of patients with ALS such as age of onset, genes obtained from single-nucleus sequencing, past brain injuries, weight, sex, and other factors with target vector being if they had ALS or not. Different genes have been linked to ALS, but not all genes (such as C9orf72) are guaranteed to present ALS, and including other genes could give better direction to combinations of genes that could cause disease. A classification example for my research could be similar to the above experiment, except the goal would be to look for the genomic-subtype of ALS (i.e. what gene seems to be the primary driver of ALS for this patient. TDP43, C9orf72, and SOD2 are the three most prominent genomic causes of ALS, and classification models could help better determine the primary driver of ALS in patients that differentially express other genes and identify what subtype of ALS future patients may have. Risks in medical models are that they are simply predictions based on current data obtained from patients we 'think' have ALS, and ALS is an unclear disease as other diseases such as frontotemporal dementia have similar neurodegenerative patterns. Furthermore, there may not be enough variation within samples as there is still very few ALS subtype datasets available.

2. a) $J_{reg}^{\beta}(w) = J + R$

gradient descent: $w_j \leftarrow w_j - \alpha$ (partial derivative of cost function)

$$w_j \leftarrow w_j - \alpha \left(\frac{\partial J_{reg}^{\beta}}{\partial w} \right) (1)$$

$$= w_j - \alpha \left(\frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) x_j^{(i)} + \sum_{j=1}^{D} \beta_j w_j \right)$$

main answer

$$\frac{1}{2} \left(2 \cdot \beta_j w_j^{2-1} \right)$$

$$= \beta_j w_j$$

If the jth feature is less important, it is best to set a greater β_j to ensure the model still trains well enough to fit the data

b) $$J = \frac{1}{2N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)})^2$$

$$= \frac{1}{2N} \sum_{i=1}^{N} \left(\sum_j w_j x_j^{(i)} + b - t^{(i)} \right)^2$$

Let b (bias term) $= 0$

$$\frac{\partial J}{\partial w_j} = \frac{1}{N} \sum_{i=1}^{N} x_j^{(i)} \left(\sum_j w_j x_j^{(i)} - t^{(i)} \right)$$

Let $\frac{\partial J}{\partial w_j} = 0$ to find minima

$$\frac{\partial J}{\partial w_j} = \frac{1}{N} \sum_{i=1}^{N} x_j^{(i)} \left(\sum_{j=1}^{D} w_j x_j^{(i)} - t^{(i)} \right).$$

Expand to obtain system of linear equations

$$\frac{\partial J}{\partial w_j} = \frac{1}{N}\sum_{j'=1}^{D}\left(\sum_{i=1}^{N} x_j^{(i)} x_{j'}^{(i)}\right) w_{j'} - \frac{1}{N}\sum_{i=1}^{N} x_j^{(i)} t^{(i)}$$

$$A_{jj'} \qquad\qquad c_j$$