Assignment Questions on Logistic Regression

Mathematics Based.

1) $$\text{log-odds} = \log\left(\frac{p}{1-p}\right) = \theta^T x.$$

Now we solve this equation for p:

① Exponentiate both sides:

$$\frac{p}{1-p} = e^{\theta^T x}$$

② Multiply both sides by 1-p:

$$p = (1-p)e^{\theta^T x}.$$

③. Expand:

$$p = e^{\theta^T x} - pe^{\theta^T x}.$$

④. Bring the p-terms together:

$$p + pe^{\theta^T x} = e^{\theta^T x}$$

$$p(1+e^{\theta^T x}) = e^{\theta^T x}.$$

⑤ Solve for p:

$$p = \frac{e^{\theta^T x}}{1+e^{\theta^T x}} = \frac{1}{1+e^{-\theta^T x}}$$

This function,

$$\sigma(z) = \frac{1}{1+e^{-z}}$$ is sigmoid (or logistic) function

2). Derivation of the logistic (log-loss) cost and convexity

<u>Hypothesis</u>

$$h_\theta(x) = \sigma(\theta^T x) = 1/1+e^{-\theta^T x}.$$

(where $\sigma(\cdot)$ is the sigmoid).

2) (continued)

Let $X$ be the $m \times n$ design matrix (rows $x^{(i)T}$), $h$ the vector with $h_i = h_\theta(x^{(i)})$, and $y$ the label vector.

Gradient

$$\nabla J(\theta) = \frac{1}{m} X^T (h - y).$$

Because $\frac{\partial}{\partial \theta} h_\theta(x) = h_\theta(x)(1 - h_\theta(x))x$ and the algebra yields the above compact form.

Hessian

$$H(\theta) = \nabla^2 J(\theta) = \frac{1}{m} X^T R X,$$

where $R$ is the $m \times m$ diagonal matrix with $R_{ii} = h_i(1-h_i)$.

Note $0 < h_i(1-h_i) \leq 1/4$.

3). Predicted Probability & class (example)

For, $x = [1,2]$ and $\theta = [0.5, -0.25]$

- Compute $\theta^T x = 0.5 \times 1 + (-0.25) \times 2 = 0.5 - 0.5 = 0.$
- Predicted probability: $h_\theta(x) = \frac{1}{1+e^{-(0)}} = 0.5.$
- Class prediction (threshold $= 0.5$): The probability is equal to the threshold, the example is generally classified as the positive class.

4).

4) Logistic Regression vs SVM.

| Aspect | Logistic Regression | Support Vector Machines |
|---|---|---|
| Assumptions | Linear relationship between features & log-odds. | Maximizes margin, can model nonlinear decision boundaries. |
| Complexity | Fast training, scalable to large datasets. | Training can be quadratic/cubic, slower for large datasets. |
| Suitability | Well-suited for interpretable, streaming, large data. | Superior for high-dimen-sional, complex, or non-linear data. |
| Output | Probabilities & interpretation | Classification, margin-based, less interpretable, Kernel trick |
| Overfitting | Regularization is standard. | Robust to outliers; needs careful tuning, especially Kernels. |

5) Real-World Applications of Logistic Regression.

Health Care: Used for predicting disease risk, such as heart disease or diabetes, by modeling the relationships between patient features and outcoming. Strengths are interpretability and robust handling of various data types; a limitation is only linear decision boundaries.

Finance: Employed for credit scoring to predict the probability of default on a loan, given the characteristics of an applicant. The main advantages are in explanatory ease, which appeals to regulatory requirements; and the limitations is in lower accuracy for nonlinear relationships within data.

6) Pipeline design for patient Classification.

Preprocessing steps:

- Data cleaning: handling missing values & outliers
- Feature scaling (standardization or normalization)

Encoding categorical features if present

- Splitting into train/test sets.

Pipeline:

Fit logistic regression to Age, BMI, Blood Pressure, features access model on validation/test set.

7). Performance Evaluation ~~Matrix~~ metrics.

Evaluate spam detection via:

- Confusion Matrix: Shows the count of true/false positives/negatives.
- Precision: Fraction of detected spam that is actually spam.
- Recall: Fraction of actual spam emails detected.
- F1-score: Harmonic mean of precision and recall for balanced assessment.

8) Failure Example and Altanative Model.

Logistic Regression assumes that there is one linear boundary separating the classes. It will therefore perform poorly on problems like XOR or concentric circles, where no straight line can clearly separate the two classes.

In such cases, nonlinear models work better; for example:

Kernel SVM (e.g., RBF kernel), which creates curved boundaries in the original space.
Decision trees / random forests, which build flexible piecewise regions that can fit complex patterns.

9) Coefficient Interpretation:-

The logistic regression coefficients indicate the change in log-odds, and thus in probability, for a one-unit increase in each feature, holding all the other features constant.

Hours studied = 0.8 : This is a positive value, indicating studying more hours increases the probability of exam success. The relatively large magnitude means study time is a strong positive predictor.

Sleephours = -0.3 : This is a negative value, so more sleep corresponds to a lower predicted probability of success. This may be because extra sleep cuts into study time. It's magnitude is smaller, making it weaker in effect as compared with study hours.

10) High precision but Low Recall (Minority class)
High precision with poor recall on the minority class
High accuracy with poor recall generally indicates that a dataset is imbalanced.
- The model learns to predict the majority class correctly, hence getting many overall predictions right,
while it misses many cases of the minority class. (low recall).

- To fix this, you can:
Data rebalancing (oversample minority, undersample majority, or use SMOTE) use class weights

such that misclassifying the minority class is penali-
-zed more. Lower the decision threshold - for
example, from 0.5 to 0.3 - to predict the minority
class more often. This can then be tuned by precision
- recall or F1-score.