Q2 Explain the general of logistic regression and the importance of f selection

- Statistical method for binary classification tasks predicting the probability that something belongs in a class. ②

- Binary classification ($y \in \{0,1\}$): $P_{Y|X,W}(1|x,w) = \sigma(w^T\phi(x)) = \frac{1}{1+e^{-w^T\phi(x)}}$ probability for $Y=1$

  $P_{Y|X,W}(0|x,w) = 1-\sigma(w^T\phi(x))$ pro

- Decision boundary: Predict 1 if $\sigma \geq 0.5$ else 0 for σ, Provides class probabilities instead hard labels

- We use MLE for probability for weights to better classify, log-likelihood good for computation $\ell(w) = \sum_{n=1}^{N} [y_n \ln\sigma(w^T\phi(x_n)) + (1-y_n)\ln(1-\sigma(w^T\phi(x_n)))]$

  Likelihood for N (i.i.d)

- Cross-Entropy Loss: $-\ell(w)$ which we minimize with gradient descent. Maximize Likelihood.
- Use gradient of log-likelihood tells direction weights increase likelihood.

---

①

- We want to find weights for each classifier, get weights (converge) so data linearly seperable.
- If data is not linearly seperable in original, then select basis functions such that is in feature space.
- Converge weights → no miss classified data

feature = f