HW-2-陈宸

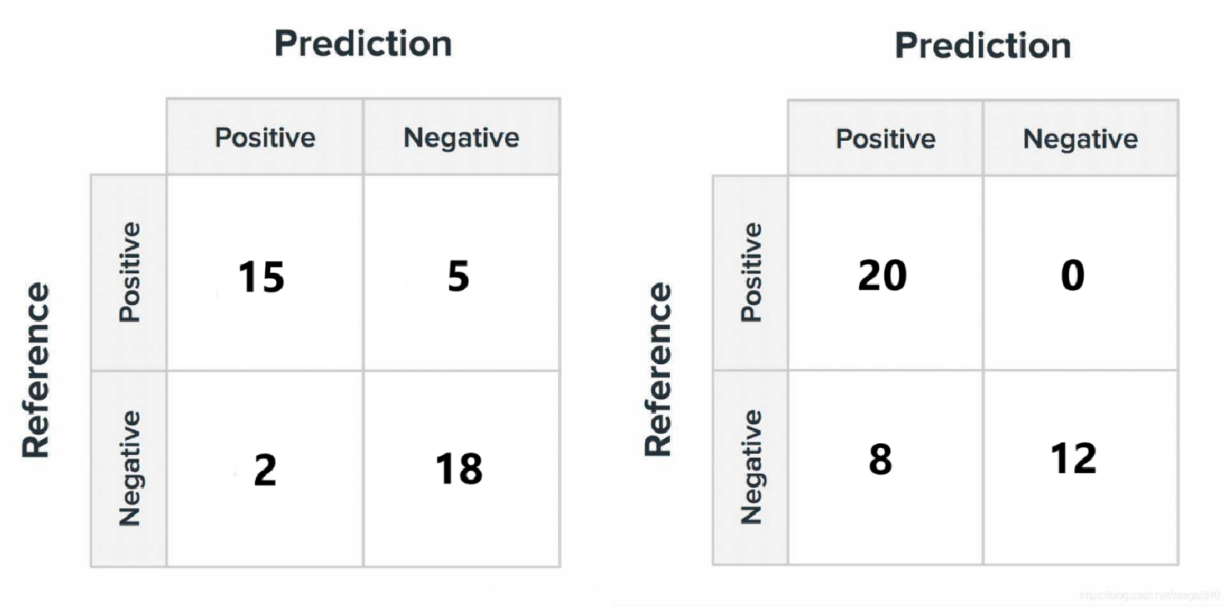
Q1. Precision and recall

sum = 40

before: TP = 15, FN = 5, FP = 2, TN = 18

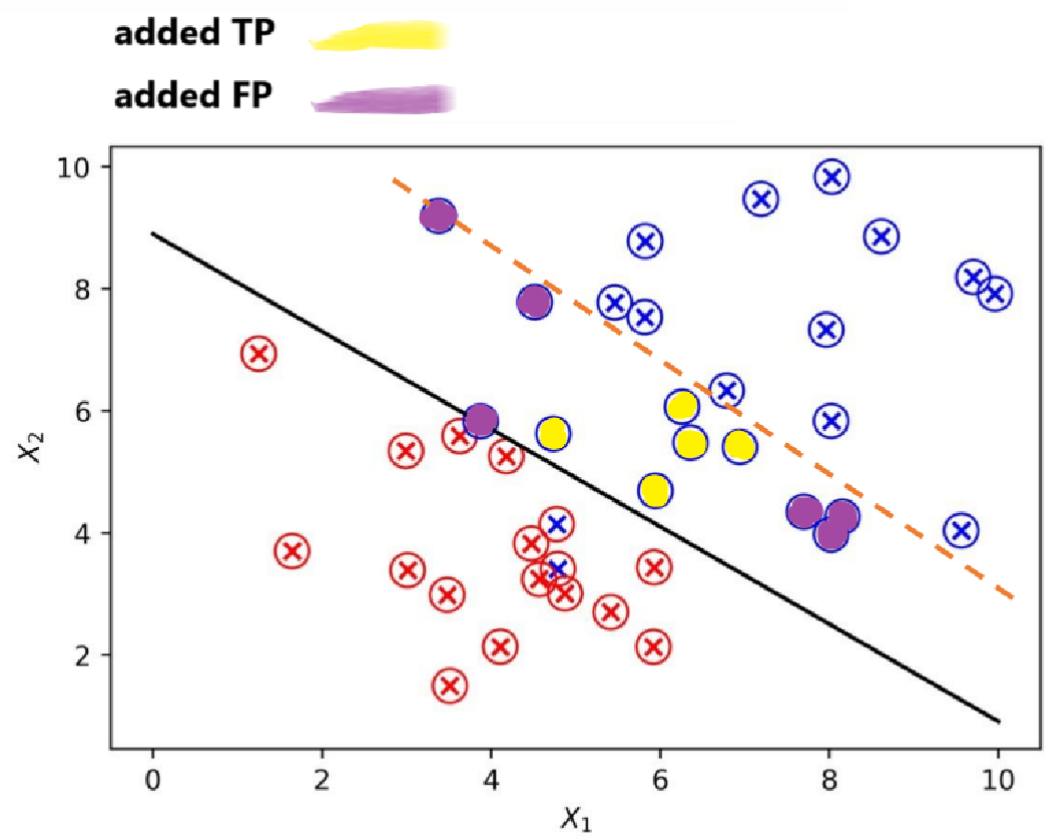
after: TP = 20, FN = 0, FP = 8, TN = 12

After changing the decision boundary, the value of TP and FP both increases.



precision = 15/15+2 = 0.88, recall = 15/15+5 = 0.75

precision = 20/20+8 = 0.71, recall = 20/20+0 = 1.00



When the value of precision increases, recall decreases, and vice versa.

Q2. Normalization

2.1

- **Normalization Dimension**: BN normalizes each feature across the mini-batch, while LN normalizes all features within a single sample.
- **Suitable Scenarios**: BN is better for large batch data processing, such as image classification tasks in CNNs; LN is better for sequential data processing, such as NLP tasks in RNNs and Transformers.
- **Training and Inference**: BN uses mini-batch statistics during training and moving average statistics during inference; LN uses the same normalization process during both training and inference.

2.2

CLIP: Uses Layer Normalization.

GPT-2: Also employs Layer Normalization.

LLAMA-3: It uses Root Mean Squared Layer Normalization, which is a variant of Layer Normalization.

2.3

CLIP

CLIP consists of two main components: a text encoder and an image encoder.

- **Text Encoder**: It uses a Transformer architecture. In embedding layer, converts input tokens into dense vectors.Then it contains transformer layers, which are multiple layers of self-attention and feed-forward neural networks. Last was the Projection Layer, it projects the final hidden states to a common embedding space.
- **Image Encoder**: It uses a Vision Transformer architecture. In Patch Embedding Layer, it divides the image into patches and embeds them.Its Transformer Layers and Projection Layers are like Text Encoder's.

GPT-2

- **Embedding Layer**: Converts input tokens into dense vectors.
- **Transformer Layers**:It consists of 12 layers of self-attention and feed-forward neural networks.
- **Self-Attention Heads**: Each layer has multiple self-attention heads.
- **Feed-Forward Neural Networks**: Each layer has a feed-forward neural network.
- **Output Layer**: Projects the final hidden states to the vocabulary size for token prediction.

LLAMA-3

Its structure consists of Embedding Layer, Self-Attention Layer, Feed-Forward Neural Network,Projection Layer, Output Layer.