**1.Activation func** 1, Leaky ReLU a, 2, Derivative: 2, Swish, a, ; b, Derivative: ;3, Mish: a, ,b, **2. Smooth Activation funcs** Swish, Mish, and GELU are advanced activation functions known for their smoothness and non-monotonic nature, allowing for better gradient propagation, crucial for learning complex patterns. In NLP, tasks involve hierarchical data processing, where understanding the context and relationships between words or phrases at different levels is key. These activation functions excel in capturing these intricate, hierarchical dependencies due to their ability to facilitate more nuanced and flexible gradient flow, enhancing model performance in deep, complex architectures. In contrast, CV tasks primarily focus on spatial hierarchies and local feature extraction, where ReLU's simplicity and efficiency in handling such patterns are sufficient. Thus, while Swish, Mish, and GELU offer advantages in the context-rich, hierarchical domain of NLP, ReLU remains effective for the spatial and structural patterns common in CV. **3.GDA** a, Assumptions: Assumes data from each class comes from a multivariate normal (Gaussian) distribution with parameters ; b, Model: 1, ; 2, ; 3, ; c, Parameters: 1, ; 2, ; 3, ; 4, . c, Decision Boundary: Quadratic in if ; Linear if .d, Essence: GDA is effective when the Gaussian assumption holds and when there is enough data to estimate the parameters accurately. It's particularly well-suited for applications where the relationship between features and classes is approximately normally distributed. **4. Dis vs Gen** Discriminative classifiers, such as logistic regression, learn the conditional probability to define the decision boundary directly, while generative classifiers, like Gaussian Discriminant Analysis, model the joint probability distribution , estimating and to infer class labels. This distinction leads to discriminative models focusing on prediction accuracy, whereas generative models provide insights into data's underlying structure and distribution. **5. Bagging** involves training multiple models (usually of the same type) on different subsets of the training data, obtained by bootstrapping (sampling with replacement). It aims to improve the stability and accuracy of machine learning algorithms by reducing variance, avoiding overfitting, and providing a way to estimate model uncertainty. Its implementation: Multiple models are trained on bootstrap samples of the training data. The final prediction is typically the average of predictions from all models (for regression) or the majority vote (for classification).