

# Deep Learning Cheat Sheet

A concise reference for core deep learning concepts, especially for building and evaluating baseline models.

---

## 1. Activation Functions

Function	Formula	Notes
ReLU	$\max(0, x)$	Fast, prevents saturation
Sigmoid	$1 / (1 + e^{-x})$	Good for binary output
Tanh	$(e^x - e^{-x}) / (e^x + e^{-x})$	Zero-centered sigmoid
GELU	Smooth variant of ReLU	Used in Transformers

---

## 2. Training Parameters

- **Batch Size:** Number of samples per training step
  - **Epoch:** One full pass through the training set
  - **Learning Rate:** Step size for parameter updates
  - **Dropout:** Randomly disables neurons to reduce overfitting
  - **Weight Decay:** L2 regularization to penalize large weights
- 

## 3. Normalization Techniques

### Batch Normalization

- Normalizes across batch dimension
- Reduces internal covariate shift

### Layer Normalization

- Normalizes across features (per sample)
  - Preferred in RNNs and Transformers
- 

## 4. Bias vs Variance

Model Type	Bias	Variance	Risk
Underfit	High	Low	Can't learn
Overfit	Low	High	Poor generalization

---

Model Type	Bias	Variance	Risk
Balanced	Low	Low	Ideal

---



---

## 5. Evaluation Metrics

- **Accuracy:** Correct predictions / Total
  - **Macro F1 Score:** Average F1 across classes equally
  - Use macro F1 when data is balanced (e.g., AG News)
- 

## 6. Attention Mechanism (Transformers)

The scaled dot-product attention formula:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$