Final Review

Perceptron

- Linear classifier: decision boundary w x + b = 0
- Update on misclassified sample:

$$w \leftarrow w + yx$$
$$b \leftarrow b + y$$

- Convergence **iff** data is linearly separable
- Limitations:
  - cannot solve XOR;
  - non-differentiable step activation

Neural Networks

- MLP: input → hidden (nonlinear) → output
- Activation functions
- Backprop: chain rule for gradients;

CNN

- Convolution:
  - parameter sharing;
  - output size via (kernel, stride, padding)
- Pooling: reduces spatial size; **no** learnable params

DL Recipe

- **Early Stopping**: stop when no improvement
- **Activation Functions**:
  - Sigmoid: vanishing gradients
  - ReLU: sparse activations, mitigates vanishing gradients
  - Leaky/Parametric ReLU: fix “dying ReLU”
- **Regularization**:
  - L2/Weight decay
  - L1: sparsity;
- **Adaptive Optimizers**:
  - Adagrad, RMSprop, Adam
- **Dropout**:
  - Train: random drop per mini-batch; implicit ensemble effect
  - Test: scale by 1-p (or use inverted dropout during training)

Unsupervised learnig

- **K-means Clustering**
  - **Goal**: Partition n data points into k clusters
  - **Limitations**:
    - Sensitive to initialization.
    - Number of clusters k must be chosen in advance.
- **Discriminative vs. Generative**
  - Note not all generative models are unsupervised!

GAN

- Two-player game: Generator $G(z)$ vs Discriminator $D(x)$
- Objective (original): $\min_G \max_D \mathbb{E}[\log D(x)] + \mathbb{E}[\log(1 - D(G(z)))]$
- Optimal $D^*(x) = \frac{p_{\text{data}}}{p_{\text{data}} + p_G}$; minimizes **JS divergence**

Autoencoders

- Encoder $x \rightarrow z$, Decoder $z \rightarrow \hat{x}$; minimize reconstruction loss (MSE / CE)
- Undercomplete AE: compression → feature learning
- AE vs. DAE vs. VAE

Transformers

- Attention mechanism
- Seq2Seq Model
- Encoder-Decoder Architecture

Checklist

- Know core formulas (perceptron update, conv shape)
- Be able to **compare** AE/VAE/GAN with one sentence each
- Practice at least one manual step for: backprop sign, conv shape, perceptron update

Likely Exam Prompts

- Compute one perceptron update on a given misclassified point
- Do CNN output shape and parameter count
- Write GAN minimax objective; explain why gradients reach G
- Etc.