

MACHINE LEARNING MUST-KNOWS

note from pranav:

- full transparency, i ripped all of this from GPT. it'll still be useful as i explore concepts i'm still unfamiliar with. it's a good reference for what **must-knows** look like in the huge field of ML. it certainly narrows my focus. maybe it'll do the same for you.

must-know ML papers

core ml & deep learning

- [attention is all you need](#) — transformers. period.
- [rethinking batch norm](#) — training stability
- [resnet](#) — residual connections, deep networks
- [alexnet](#) — deep learning renaissance
- [u-net](#) — segmentation architecture
- [dropout](#) — regularization via randomness
- [understanding deep learning requires rethinking generalization](#) — why deep nets generalize at all

representation learning

- [simclr](#) — contrastive pretraining
- [byol](#) — bootstrap your own latent
- [dino](#) — self-distillation without labels
- [bert](#) — masked language modeling
- [word2vec](#) — OG embeddings

generative models

- [vae](#) — probabilistic generative modeling
- [gan](#) — adversarial training
- [diffusion models beat gans](#) — DMs take over
- [stable diffusion](#) — open-source image generation
- [glide](#) — text-conditioned diffusion

optimization & theory

- [adam](#) — optimizer that just works
 - [lottery ticket hypothesis](#) — pruning magic
 - [information bottleneck](#) — theory of representation
 - [double descent](#) — more params, less error... sometimes
-

must-know concepts

don't just memorize — *grasp the intuition* behind these

optimization

- gradient descent
- loss functions (mse, cross-entropy, triplet)
- weight decay vs l2 regularization
- dropout
- adam vs sgd vs rmsprop
- learning rate schedules (cosine, step decay)

- batch size effects

architecture

- attention (self vs cross)
- positional encoding (sinusoidal vs learned)
- residual connections
- batch norm vs layer norm
- multi-head attention
- feedforward blocks
- convolution basics (stride, padding, dilation)

generative modeling

- vae
- gan
- diffusion models
- latent space interpolation
- classifier-free guidance
- denoising score matching
- noise scheduling

embeddings & representation

- word2vec, glove

- simclr, byol, dino
- contrastive loss
- cosine similarity
- infoNCE
- anchor-positive-negative

theory

- backpropagation
- svd & pca
- kl divergence
- information bottleneck
- overfitting & underfitting
- overparameterization
- double descent
- expressivity of deep networks

resources

you don't need 100 courses — just these few:

video / visual

- [3blue1brown NN playlist](#)
- [StatQuest \(Josh Starmer\)](#)
- [Jay Alammar blogs](#)

- [AI Coffee Break w/ Letitia](#)
- [Yannic Kilcher](#) — deep paper reviews

hands-on

- [karpathy's micrograd](#) — write backprop from scratch
- [fastai course](#) — project-first ML
- [pytorch tutorials](#)
- [google's machine learning crash course](#)

reference & reading

- [PapersWithCode](#)
- [arxiv-sanity](#)
- [distill.pub](#) — visual + rigorous
- [lilianweng.github.io](#) — mini research posts
- [cs231n \(vision\)](#)
- [cs224n \(nlp\)](#)
- [deep learning book](#) — by goodfellow et al.