1. Αξιολόγηση & Μετρικές

Confusion Matrix: Acc=$\frac{TP+TN}{N}$ | Prec=$\frac{TP}{TP+FP}$ Rec=$\frac{TP}{TP+FN}$ | Spec=$\frac{TN}{TN+FP}$ | $F1 = \frac{2PR}{P+R}$.
Medical Bayes: $P(Diseas|+) = \frac{Sens \cdot Prev}{Sens \cdot Prev + (1-Spec)(1-Prev)}$.
Errors: $MSE = \frac{1}{N}\sum(y-\hat{y})^2$. $R^2 = 1 - \frac{RSS}{TSS}$.
Bias-Var: $Err = Bias^2 + Var + Noise$.
• High Bias → Underfit (Simple). • High Var → Overfit (Complex).
ROC/AUC: Plot TPR vs FPR. Random=0.5, Perfect=1.
Cross-Val: k-Fold (Low Bias/Var est), Hold-out (Fast).

2. Probabilities & Info

Bayes: $P(h|d) \propto P(d|h)P(h)$ (Post Lik Prior).
Bayes Risk: Choose action w/ min exp. loss.
Gauss: $\mathcal{N} \sim |\Sigma|^{-1/2} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$.
MLE: $\hat{\mu} = \bar{x}, \hat{\sigma}^2 = \frac{1}{N}\sum(x-\mu)^2$ (Biased). Unbiased: divide $N-1$.
Entropy: $H(P) = -\sum P \log P$. Max at uniform.
KL Div: $D_{KL} = \sum P \ln \frac{P}{Q}$. **CrossEnt**: $H(P,Q) = H(P) + D_{KL}$.
MLE: $\max \sum \ln P(x|\theta)$. $\nabla \ln L = 0$.

3. Γραμμικά Μοντέλα

LinReg: $w = (X^TX)^{-1}X^Ty$. ($O(N^3)$).
LogReg: $P = \sigma(w^Tx) = \frac{1}{1+e^{-z}}$. Convex Loss.
$\sigma' = \sigma(1-\sigma)$. Update: $w \leftarrow w + \eta(y-\hat{y})x$.
LDA: Max Sep $J(w) = \frac{w^TS_Bw}{w^TS_Ww}$.
$S_B = (m_2-m_1)(m_2-m_1)^T$. $S_W = \sum S_i$.
Opt: $w \propto S_W^{-1}(m_2-m_1)$. Assumes equal Σ.
Geometry: Hyperplane $w^Tx + w_0 = 0$.
Dist $r = y(x)/||w||$. $w \perp$ surface.
Regularization (Shrinkage):
L1 (Lasso): $\lambda|w|$ (Sparsity/Selection).
L2 (Ridge): λw^2 (Small weights).

4. SVM & Kernels

Primal: $\min \frac{1}{2}||w||^2 + C\sum\xi_i$ s.t. $y_i(w^Tx_i + b) \geq 1-\xi_i$.
Convex: Global minimum guaranteed.
Support Vectors: Points on margin ($y(\cdot) = 1$) or errors. Only these affect w.
Margin: $2/||w||$. Larger C → Harder (Sm margin).
Slacks ξ: 0 (correct), $0 < \xi < 1$ (margin vio), $\xi > 1$ (error).
Kernel Trick: $x^Tz \to K(x,z)$. Implicit high dim.
• Poly: $(x^Tz + c)^d$. Dim $\approx d$-order terms.
• RBF: $e^{-\gamma||x-z||^2}$. Dim ∞. (Taylor exp).
Geometry: Max margin boundary is $\perp$ bisector of closest points from classes.

5. Neural Networks

Non-Convex: Local optima possible.
Unit: $z = \sum w_ix_i + b \to a = g(z)$.
Activ: ReLU $\max(0,z)$ (Fast, no vanish grad).
Sigmoid (0,1), Tanh (-1,1) (Sat at limits).
Softmax: $y_k = e^{z_k}/\sum e^{z_j}$. For Multi-class.
Forward: $Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}, A^{[l]} = g(Z^{[l]})$.
Backprop: Chain rule $\frac{\partial J}{\partial W}$.
$\delta^L = (a^L - y)$, $\delta^l = (W^{l+1})^T\delta^{l+1} \cdot g'(z^l)$.
$\partial J/\partial W^l = \delta^l(a^{l-1})^T$.
Params: FC: $(N_{in}+1)N_{out}$.
CNN: $Out = \lfloor\frac{W-K+2P}{S}\rfloor + 1$.
Params: $F \times F \times C_{in} \times C_{out} + C_{out}$.
Optim: SGD (Noisy), Batch (Slow), Mini-batch.
Momentum (Velocity), Adam (Adapt LR).

6. Ensembles

Bagging: Bootstrap (replace) + Parallel Models.
Reduces **Variance**. Ex: Random Forest.
Random Forest: Bagging + Feature subset ($\sqrt{p}$).
Decorrelates trees -> Better reduction of var.
Boosting: Sequential. Fix prev errors.
Reduces **Bias** (and Var). Ex: AdaBoost, XGB.
AdaBoost: Weights $\alpha_t = \frac{1}{2}\ln\frac{1-\epsilon}{\epsilon}$.
Data weights $D_{t+1} \propto D_te^{-\alpha y\hat{y}}$ (Focus Hard).
Trees: Split: Max InfoGain ($H_{pre} - H_{post}$).
Entropy $-\sum p\log p$, Gini $1-\sum p^2$.

7. Unsupervised & PCA

PCA: Max Variance directions (Ortho).
1. Center $(x-\mu)$. 2. Cov $\Sigma = \frac{1}{N}X^TX$. 3. Eigendecomp $U\Lambda U^T$.
Proj $z = U_k^Tx$. Var ratio $\sum_1^k\lambda_i/\sum\lambda_{all}$.
Included '1' feature doesn't change PCA (const var=0).
K-Means: Iterate: Assign closest, Update centroids.
Converges finite steps. Spherical clusters. Sensitive to K/Init.
GMM: Soft K-means. Probabilistic (π, μ, Σ).
EM Algo: Max Likelihood (local max).
E-step: Calc responsibilities γ. M-step: Update params.
DBSCAN: Density. core/border/noise. Args: eps, minPts.
Hierarchical: Agglomerative.
Linkage: Single (min dist), Complete (max), Avg.
Silhouette: $\frac{b-a}{\max(a,b)}$. +1 Good, 0 Border, -1 Wrong.

8. KNN & Dim Reduction

KNN: Lazy. Distances (L_2).
Small K → High Var (Noise). Large k → High Bias (Smooth).
Curse of Dim: Vol grows exp. Points equidistant.
Need exp data. Solutions: PCA, Feature Sel.

9. True/False Exam Bible (1/2)

-- SVM Objective: Convex Quatratic (Global Min).
-- NN Objective: Non-Convex (Local Min).
-- EM Algo: Maximizes Likelihood (Local Max).
-- K-Means: Loss decreases monotonically. Converges.
-- RBF Feature Space: Infinite dimensional.
-- PCA: Adding const feature -> No change.
-- Perceptron: Oscillates if not lin sep.
-- Hard Margin SVM: Requires lin sep.
-- Bootstrapping: Sampling WITH replacement.

10. True/False Exam Bible (2/2)

-- L1 Reg: Sparsity (Feat Sel). L2: Small Weights.
-- Bias-Var: 1-NN (High Var), K-NN (High Bias).
-- Kernel Trick: w is lin comb of data (Representer).
-- Softmax: Used for Multi-class Output.
-- Linear Activ in Hidden: Collapse to linear model.
-- Generative: Naive Bayes, GMM, LDA.
-- Discriminative: LogReg, SVM, NN, KNN.
-- Decision Trees: Pruning reduces overfitting.

11. Quick Math/Formulas

Logarithms: $\ln(xy) = \ln x + \ln y$, $\ln(e^x) = x$.
Derivatives: $\frac{d}{dx}\sigma(x) = \sigma(1-\sigma)$.
Matrix: $(AB)^T = B^TA^T$. $\nabla_x w^Tx = w$.
$\nabla_x x^TAx = (A+A^T)x$.
Distances: L_1 Manhattan $\sum|x|$, L_2 Euclidean $\sqrt{\sum x^2}$.
Eigen: $Av = \lambda v$. $\det(A-\lambda I) = 0$.
Normal: 68% (1σ), 95% (2σ), 99.7% (3σ).

12. Algorithm Summary

Naive Bayes: $P(y|x) \propto P(y)\prod P(x_i|y)$. Fast, High Bias.
LogReg: Linear boundary. Intepretable. P-values.
SVM: Max Margin. Robust. Kernels → Non-Lin.
DT: Non-parametric. Interp. Orthogonal splits.
RF/GBM: Black-box. High Acc. Robust.
NN: Univ approx. Needs data/compute. Non-convex.
K-Means: Simple. Scalable. Convex clusters only.