

40 Models for Mapping Inputs → Outputs

A compact field-guide collecting **formulations**, where they're used, **history**, **use cases**, and **what they comprise**—all in one place.

Notation (lightweight): inputs i (or sequence $i_{1:T}$), outputs o (or $o_{1:T}$); dataset $\mathcal{D} = \{(i_k, o_k)\}$. Vectors are bold implicitly; noise terms ϵ, w, v as context requires.

1) Deterministic function (static map)

Formulation. $o_k = f(i_k)$. - **Where:** rule systems, ETL transforms, calculators.

- **History:** classical baseline from mathematical analysis.
- **Use cases:** scoring rules, lookup tables, calibrations.
- **Comprises:** explicit rule set or closed-form expression; no estimation.

2) Linear / affine model

Formulation. $o_k = W i_k + b$ (vector) or $o_k = w^\top i_k + b$ (scalar). - **Where:** baselines across science/engineering.

- **History:** Gauss/Legendre, early 1800s (least squares).
- **Use cases:** prediction, forecasting, control.
- **Comprises:** weights, bias; least squares or regularized fitting.

3) Basis expansion (poly/splines/Fourier)

Formulation. $o_k = \sum_j \alpha_j \phi_j(i_k)$. - **Where:** statistics, signal processing, curve fitting.

- **History:** Fourier (19th c.); splines (Schoenberg, 1946).
- **Use cases:** nonlinearity with interpretability, seasonality, smooth trends.
- **Comprises:** dictionary of basis functions; linear coefficients; regularization.

4) k-Nearest Neighbors (kNN)

Formulation. $\hat{o}(i) = \frac{1}{k} \sum_{j \in \mathcal{N}_k(i)} o_j$. - **Where:** low-dimensional data; quick baselines.

- **History:** Fix & Hodges (1951).
- **Use cases:** classification/regression, imputation.
- **Comprises:** distance metric, k , tie-breaking; no training phase beyond indexing.

5) Kernel regression (Nadaraya–Watson)

Formulation. $\hat{o}(i) = \frac{\sum_j K(i, i_j) o_j}{\sum_j K(i, i_j)}$. - **Where:** smoothing/LOESS-style fits.

- **History:** 1964.

- **Use cases:** calibration curves, local averaging.
- **Comprises:** kernel choice and bandwidth; sometimes local polynomial variants.

6) Logistic / softmax models

Formulation. Binary: $\Pr(o = 1 \mid i) = \sigma(w^\top i + b)$. Multiclass: $\Pr(o = c \mid i) = \frac{e^{w_c^\top i + b_c}}{\sum_{c'} e^{w_{c'}^\top i + b_{c'}}$. - **Where:**

classification, propensity modeling.

- **History:** logistic curve (19th c.), logistic regression (1950s).
- **Use cases:** credit risk, medical diagnosis, click-through.
- **Comprises:** linear scores + link; cross-entropy; regularization.

7) Decision trees / ensembles (RF/GBDT)

Formulation. Piecewise model $o = \sum_{\ell} \theta_{\ell} \mathbf{1}[i \in R_{\ell}]$; ensembles average/vote.

- **Where:** tabular data.
- **History:** CART (1984); Random Forests (2001); Gradient Boosting (2001+).
- **Use cases:** ranking, churn, credit scoring, feature importance.
- **Comprises:** splits, impurity metrics; bagging/boosting, shrinkage, depth/trees.

8) Gaussian processes (GP)

Formulation. $f \sim \mathcal{GP}(m, k)$, $o = f(i) + \epsilon$. Predictive mean $\mu_* = k_*^\top (K + \sigma^2 I)^{-1} y$; variance $\sigma_*^2 = k(i_*, i_*) - k_*^\top (K + \sigma^2 I)^{-1} k_*$. - **Where:** small/medium data with uncertainty needs.

- **History:** Kriging (1930s); modern GPs (1990s–2000s).
- **Use cases:** Bayesian optimization, spatial stats, calibration.
- **Comprises:** kernel, mean function; exact or sparse inference.

9) Mixture of experts (MoE)

Formulation. $p(o \mid i) = \sum_{m=1}^M \pi_m(i) p_m(o \mid i)$. - **Where:** heterogeneous regimes, large models (routing).

- **History:** Jacobs/Jordan (1991); modern sparse MoE for scaling.
- **Use cases:** multimodal outputs, domain specialization.
- **Comprises:** experts (NNs/GLMs), gating network, mixture training.

10) Bayesian likelihood model

Formulation. Prior $p(\theta)$, likelihood $p(o \mid i, \theta)$; posterior $p(\theta \mid \mathcal{D}) \propto p(\theta) \prod_k p(o_k \mid i_k, \theta)$. - **Where:** principled uncertainty, hierarchical pooling.

- **History:** Bayes (1763); MCMC/VI (1990s+).
- **Use cases:** A/B tests, meta-analysis, small-data regimes.
- **Comprises:** prior, likelihood family, inference (MCMC/VI), posterior predictive.

11) Noisy memoryless channel

Formulation. $p(o | i)$; analyze capacity $\max_{p(i)} I(I; O)$. - **Where:** communications, error modeling (OCR/ASR).

- **History:** Shannon (1948).
- **Use cases:** coding, denoising, robust decoding.
- **Comprises:** channel law (e.g., BSC, AWGN), information measures.

12) Causal SEM / DAGs

Formulation. For variables X_j : $X_j = f_j(\text{Pa}(X_j), U_j)$, acyclic graph; interventions via $do(\cdot)$. - **Where:** policy, epidemiology, economics.

- **History:** Wright (1920s), Pearl (1990s).
- **Use cases:** counterfactuals, mediation, uplift.
- **Comprises:** graph, structural equations, exogenous noise, identification.

13) LTI / convolutional system

Formulation. $o_t = (h * i)_t = \sum_{\tau \geq 0} h_\tau i_{t-\tau}$; transfer $H(z) = \sum_{\tau} h_\tau z^{-\tau}$. - **Where:** control, DSP, audio, imaging.

- **History:** 1930s–1950s systems theory.
- **Use cases:** filters, deconvolution, equalization.
- **Comprises:** impulse response, stability, frequency response.

14) Linear state-space (Kalman)

Formulation. $x_{t+1} = Ax_t + Bi_t + w_t$; $o_t = Cx_t + Di_t + v_t$.

- **Where:** tracking, navigation, sensor fusion.
- **History:** Kalman (1960).
- **Use cases:** robotics, aerospace, econometrics nowcasting.
- **Comprises:** system matrices, noise covariances; Kalman filter/smoothing.

15) ARIMAX / dynamic regression

Formulation. $o_t = \sum_{j=1}^p \phi_j o_{t-j} + \sum_{m=0}^q \beta_m i_{t-m} + \epsilon_t$ (with differencing/MA terms as needed). - **Where:** forecasting with exogenous drivers.

- **History:** Box–Jenkins (1970).
- **Use cases:** demand, price, traffic forecasting.
- **Comprises:** AR/MA orders, differencing, exogenous regressors.

16) Input–Output HMM (IO-HMM)

Formulation. $p(z_{t+1} | z_t, i_t)$, $p(o_t | z_t, i_t)$; latent z_t . - **Where:** controlled/semi-Markov sequences.

- **History:** 1990s extensions of HMMs.

- **Use cases:** dialogue systems, bio/finance regimes.
- **Comprises:** state set, input-conditioned transitions/emissions; EM/inference.

17) RNN / LSTM / GRU

Formulation. $h_t = \phi(W_{ih}i_t + W_{hh}h_{t-1} + b)$, $o_t = W_{ho}h_t + c$. LSTM/GRU add gates.

- **Where:** sequential ML.
- **History:** RNN (1980s), LSTM (1997), GRU (2014).
- **Use cases:** language, sensor streams, anomaly detection.
- **Comprises:** recurrent cell, hidden state, optimizer, regularization.

18) Transformer (seq2seq/attention)

Formulation. $\text{Attn}(Q, K, V) = \text{softmax}(QK^\top / \sqrt{d})V$; encoder/decoder stacks with positional encodings.

- **Where:** NLP, vision, time-series.
- **History:** 2017 onward.
- **Use cases:** translation, summarization, forecasting, retrieval.
- **Comprises:** self/cross-attention blocks, MLPs, normalization, residuals.

19) Contextual bandit

Formulation. Choose action $a_t \in \mathcal{A}$ from context x_t ; observe reward $r_t \sim p(r \mid x_t, a_t)$; learn policy $\pi(a \mid x)$. - **Where:** online decisioning.

- **History:** 2000s; LinUCB/Thompson sampling variants.
- **Use cases:** recommendations, ads, UI optimization.
- **Comprises:** exploration (UCB/TS), reward model, regret analysis.

20) MDP / Reinforcement learning

Formulation. $p(s_{t+1} \mid s_t, a_t)$; objective $\max_{\pi} \mathbb{E}[\sum_t \gamma^t r_t]$. - **Where:** control, operations, games.

- **History:** Bellman (1950s); Sutton & Barto (1998).
- **Use cases:** robotics, inventory, scheduling, games.
- **Comprises:** state, action, reward, dynamics; value functions/policies.

21) Generalized linear model (GLM)

Formulation. $g(\mathbb{E}[o \mid i]) = \beta_0 + \beta^\top i$ with exponential-family likelihood. - **Where:** classical stats/actuarial.

- **History:** Nelder & Wedderburn (1972).
- **Use cases:** counts (Poisson), rates, insurance pricing.
- **Comprises:** link function, linear predictor, dispersion; MLE/IRLS.

22) Generalized additive model (GAM)

Formulation. $g(\mathbb{E}[o | i]) = \alpha + \sum_j s_j(i_j)$ with smoothers. - **Where:** interpretable nonlinear modeling.

- **History:** Hastie & Tibshirani (1986).
- **Use cases:** risk scores, uplift, partial dependence.
- **Comprises:** spline bases, penalties, backfitting.

23) Quantile regression

Formulation. $\min_{\beta} \sum_k \rho_{\tau}(o_k - \beta^{\top} i_k)$, $\rho_{\tau}(u) = \max\{\tau u, (\tau - 1)u\}$. - **Where:** distributional/risk estimates.

- **History:** Koenker & Bassett (1978).
- **Use cases:** VaR, service levels, asymmetric costs.
- **Comprises:** pinball loss, per-quantile fits or monotone joint fits.

24) SVM / SVR

Formulation (classification). $\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum \xi_k$ s.t. $y_k(w^{\top} \phi(i_k) + b) \geq 1 - \xi_k$.

Formulation (regression). ϵ -insensitive SVR with slack.

- **Where:** medium-size tabular/text.
- **History:** Vapnik (1990s).
- **Use cases:** margin-robust classification, robust regression.
- **Comprises:** kernel, support vectors, C/ϵ hyperparams.

25) MARS

Formulation. $o \approx \sum_m c_m B_m(i)$ where B_m are data-selected hinge bases.

- **Where:** response-surface modeling.
- **History:** Friedman (1991).
- **Use cases:** nonlinear tabular prediction with interpretability.
- **Comprises:** forward add / backward prune; GCV.

26) Isotonic regression

Formulation. Monotone f minimizing $\sum_k (o_k - f(i_k))^2$. - **Where:** monotone relationships, calibration.

- **History:** 1950s; pool-adjacent-violators algorithm (PAVA).
- **Use cases:** dose-response, score calibration.
- **Comprises:** monotonicity constraints; piecewise constant/linear fit.

27) Penalized linear models (ridge/lasso/elastic net)

Formulation. Ridge: $\min_{\beta} \|o - I\beta\|_2^2 + \lambda \|\beta\|_2^2$. Lasso: $+\lambda \|\beta\|_1$. EN: mix.

- **Where:** high-dimensional regression.
- **History:** ridge (1970), lasso (1996).

- **Use cases:** shrinkage, feature selection, stability.
- **Comprises:** penalty type, path algorithms (LARS), CV.

28) Reduced-rank regression (RRR)

Formulation. Multi-output: $O \approx IB$ with $\text{rank}(B) \leq r$ (factorization $B = UV^\top$).

- **Where:** multivariate outputs.
- **History:** 1950s–1970s multivariate stats.
- **Use cases:** dimension-reduced mappings, CCA-like tasks.
- **Comprises:** low-rank constraint; SVD-based solutions.

29) Mixture density network (MDN)

Formulation. NN outputs $\{\pi_m(i), \mu_m(i), \Sigma_m(i)\}$; $p(o | i) = \sum_m \pi_m \mathcal{N}(o; \mu_m, \Sigma_m)$. - **Where:** multimodal continuous targets.

- **History:** Bishop (1994).
- **Use cases:** inverse kinematics, ambiguous regressions.
- **Comprises:** neural backbone; mixture NLL training.

30) Conditional normalizing flow (cNF)

Formulation. Invertible $z = f_\theta(o; i)$; $\log p(o | i) = \log p_Z(z) + \log \left| \det \frac{\partial f_\theta}{\partial o} \right|$. - **Where:** flexible conditional densities.

- **History:** 2015–2018 (NICE/RealNVP/Glow).
- **Use cases:** generative regression, SBI, simulation surrogates.
- **Comprises:** invertible blocks, base density, exact likelihood.

31) Copula-based regression

Formulation. $F_{O,I}(o, i) = C(F_O(o), F_I(i))$; derive $p(o | i)$ via conditional copula.

- **Where:** dependence beyond correlation.
- **History:** Sklar (1959); applied widely in finance/insurance.
- **Use cases:** joint risk, tail dependence.
- **Comprises:** marginal models + copula family (Gaussian/t/Archimedean).

32) Cox proportional hazards (survival)

Formulation. $\lambda(t | i) = \lambda_0(t) \exp(\beta^\top i)$; partial likelihood for β . - **Where:** biostatistics, churn/time-to-event.

- **History:** Cox (1972).
- **Use cases:** retention, reliability.
- **Comprises:** baseline hazard, proportional effect, censoring handling.

33) Ordinal regression (proportional odds)

Formulation. $\Pr(o \leq c \mid i) = \sigma(\theta_c - \beta^\top i)$ for ordered classes c . - **Where:** ratings, grades, stages.

- **History:** McCullagh (1980).
- **Use cases:** Likert outcomes, severity scales.
- **Comprises:** cutpoints θ_c , shared slope(s), logit/probit link.

34) Zero-inflated / hurdle count models

Formulation. Mixture with inflation at zero: $\Pr(o = 0 \mid i) = \pi(i) + [1 - \pi(i)]f(0 \mid i)$, else $o \sim f(\cdot \mid i)$ (Poisson/NB). - **Where:** sparse counts.

- **History:** 1990s.
- **Use cases:** claims, defects, clicks.
- **Comprises:** zero-process + count component; logit + log link.

35) Conditional random field (CRF)

Formulation. $p(o_{1:T} \mid i_{1:T}) \propto \exp\left(\sum_t \theta^\top f(o_{t-1}, o_t, i_{1:T}, t)\right)$. - **Where:** structured prediction.

- **History:** Lafferty et al. (2001).
- **Use cases:** NER, segmentation, labeling.
- **Comprises:** feature potentials, global normalization; DP inference.

36) Energy-based model (EBM)

Formulation. Define energy $E_\theta(o, i)$; $p(o \mid i) \propto e^{-E_\theta(o, i)}$. - **Where:** generative modeling, anomaly detection.

- **History:** Boltzmann machines (1980s); modern EBMs (2000s+).
- **Use cases:** denoising, score-based methods, retrieval.
- **Comprises:** energy network; contrastive/score training; sampling.

37) Volterra series

Formulation. $o_t = \sum_\tau h_1(\tau)i_{t-\tau} + \sum_{\tau_1, \tau_2} h_2(\tau_1, \tau_2)i_{t-\tau_1}i_{t-\tau_2} + \dots$. - **Where:** weakly nonlinear systems.

- **History:** early 20th c.
- **Use cases:** RF, biomedical devices, loudspeaker modeling.
- **Comprises:** kernels of increasing order; truncation/regularization.

38) Hammerstein-Wiener block models

Formulation. Static nonlinearity \rightarrow LTI (Hammerstein) or LTI \rightarrow nonlinearity (Wiener): $o_t = (H * g(i))_t$ or $o_t = g((H * i)_t)$. - **Where:** control and system ID.

- **History:** 1930s-1950s.
- **Use cases:** actuator/sensor nonlinearities, saturation.
- **Comprises:** choice/order of blocks; identification per block.

39) NARX (nonlinear AR with exogenous input)

Formulation. $o_t = F(o_{t-1:t-p}, i_{t:t-q}) + \epsilon_t$. - **Where:** nonlinear time-series with drivers.

- **History:** Billings (1980s).
- **Use cases:** industrial processes, macro, energy load.
- **Comprises:** lag selection, nonlinear F (e.g., NN), regularization.

40) Nonlinear SDE state-space

Formulation. $dx_t = f(x_t, i_t) dt + G dW_t$, measurement $o_t = h(x_t) + v_t$; discretize for inference. -

Where: stochastic dynamics.

- **History:** Itô calculus (1940s); EKF/UKF/particle filters (1960s–1990s).
- **Use cases:** finance, biology, target tracking.
- **Comprises:** drift/diffusion, measurement model, Bayesian filters.

Quick selection guidance

- **Tabular, little preprocessing:** Trees/GBDT (7), GLM/GAM (21–22), penalized linear (27).
- **Uncertainty needed with small N:** GP (8), Bayesian (10).
- **Sequential with control:** Kalman/State-space (14), ARIMAX (15), IO-HMM (16), NARX (39).
- **Long sequences/text:** RNN/LSTM/GRU (17), Transformer (18).
- **Multimodal outputs:** MoE (9), MDN (29), cNF (30).
- **Causal questions:** SEM/DAGs (12).
- **Counts/zero-heavy:** GLM (Poisson/NB) (21), zero-inflated/hurdle (34).
- **Ordered labels / survival:** Ordinal (33), Cox (32).

Want this exported to PDF/Word or trimmed to a one-pager cheat sheet? I can generate that too.