

Modelo	Modelo Matemático	Función de Costo	Estrategia de Optimización	Relación Teórica (Punto 1)	Penalización / Control	Escalabilidad	Explicación conceptual
SGDR Regressor	$t = \Phi w + \eta$	Igual a Ridge / Lasso pero por lotes.	Gradiente estocástico descendiente	Versión computacional de Ridge / MAP.	Igual que el modelo base (L_1, L_2).	Muy Alta. Entrena por minilotes, sin inversión de matrices.	Minimiza el error cuadrático o regularizado actualizando los pesos paso a paso en cada iteración, permitiendo escalar a millones de datos.
Bayesian Ridge	$t = \Phi w + \eta$, $p(w) = N(0, \Sigma_w)$	Maximiza $p(w)$	t posterior	Inferencia Bayesiana analítica	Bayesiano lineal gaussiano	Prior gaussiano controla la magnitud de w .	Media. Inversión de $P \times P$, $O(P^3)$.
Kernel Ridge	$t = f(x) + \eta$, $f(x) = \phi(x)^T w$	$\min_w \frac{1}{2} a^T (K + \lambda I) a - t^T a$	Solución dual	Regresión Rígida Kernel	Penaliza norma en espacio de características	Baja Inversión de $K(N \times N)$, $O(N^3)$	Busca minimizar el error cuadrático en un espacio transformado no lineal, donde los datos se vuelven más separables.
Gaussian Process Regressor	$f(x) \sim GP(0, K(x, x'))$	Maximiza log-ver. implit marginal	Inversión de K	Proceso Gaussiano	Kernel controla suavidad y Varianza	Muy baja. Inversión $O(N^3)$, memoria $O(N^2)$.	No minimiza una función puntual, sino que modela una distribución completa de funciones posibles, buscando la que mejor explique los datos y su incertidumbre.
SVR	$f(x) = w^T \phi(x) + b$	$\frac{1}{2} \ w\ ^2 + C \sum_i \xi_i$, s.a.	$t_i - f(x_i)$	Mínimos cuadrados Regularizados	Optimización cuadrática	Modelo Kernelizado no probabilístico.	Penaliza norma de w y errores $> \epsilon$

Modelo	Modelo matemático	Función de costo	Estrategia de optimización	Relación teórica (punto 1)	Penalización Control	Escalabilidad	Explicación Conceptual
Linear Regression	$t = \Phi w + \eta$	$\min_w \ t - \Phi w\ _2^2$	Solución analítica: $w = (\Phi^T \Phi)^{-1} \Phi^T t$	Mínimos cuadrados y Máxima Verosimilitud	No penaliza	Alta. Inversión de matriz $P \times P$, $O(P^3)$	Busca minimizar la suma de errores al cuadrado entre las predicciones y los datos reales. Ajusta la línea que mejor se adapta globalmente.
Ridge	$t = \Phi w + \eta$	$\min_w \ t - \Phi w\ _2^2 + \lambda \ w\ _2^2$	Análítica	MAP con prior gaussiano / Regularizado L_2	Penaliza grandes coeficientes (L_2).	Alta. Igual a Linear Regression, pero más estable numéricamente	Busca minimizar el error cuadrático y al mismo tiempo mantener los pesos pequeños, evitando que el modelo dependa excesivamente de variables individuales.
Lasso	$t = \Phi w + \eta$	$\min_w \ t - \Phi w\ _2^2 + \lambda \ w\ _1$	Coordinate Descent / LARS	MAP con prior Laplaciano (L_1)	Penaliza con magnitud absoluta (L_1)	Alta. Iterativo $O(N \cdot P)$	Minimiza el error cuadrático pero además reduce a cero los coeficientes menos útiles, seleccionando automáticamente las variables más relevantes.
ElasticNet	$t = \Phi w + \eta$	$\min_w \ t - \Phi w\ _2^2 + \lambda_1 \ w\ _1 + \lambda_2 \ w\ _2^2$	Coordinate Descent	Combinación Ridge + Lasso	Penalización mixta ($L_1 + L_2$)	Alta. Iterativo y eficiente	Busca minimizar el error cuadrático combinando dos restricciones: que los pesos sean pequeños (Ridge) y que muchos sean ceros (Lasso), equilibrando simplicidad y robustez.

Modelo	Modelo Matemático	Función de costo	Estrategia de Optimización	Relación teórica (punto 1)	Penalización / Control	Escalabilidad	Explicación conceptual
Random Forest Regressor	$f(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$	Minimiza error cuadrático promedio	Bagging + árboles	No paramétrico / Error cuadrático	Control estructural (número y profundidad de árboles).	Alta. Paralelizable $O(B \cdot N \cdot \log N)$	Minimiza la varianza de predicción al combinar múltiples árboles que aprenden distintos patrones del mismo conjunto.
Gradient Boosting XGBoost	$f(x) = \sum_{m=1}^M \gamma_m h_m(x)$	Minimiza error cuadrático mediante boosting	Gradiente aditivo secuencial	Extensión iterativa del error cuadrático	Controla complejidad con tasa de aprendizaje y profundidad	Alta. Distribuido $O(M \cdot N \cdot \log N)$	Minimiza el error residual de los modelos anteriores agregando árboles que corrigieron progresivamente los errores previos.