Common Regression Methods

Name	Formula	Definition	Significance
Ordinary Least Squares (OLS)	$\min_\beta \sum_{i=1}^n (y_i - X_i\beta)^2$	Minimizes the sum of squared residuals between observed and predicted values	Provides unbiased, efficient estimates under classical assumptions; foundation for many statistical models. **Closed-form:** Yes ($\beta = (X^TX)^{-1}X^Ty$); **Cost:** $O(nd^2 + d^3)$
Ridge Regression	$\min_\beta \sum_{i=1}^n (y_i - X_i\beta)^2 + \lambda\|\beta\|_2^2$	OLS with ℓ_2 penalty on coefficients	Shrinks coefficients to reduce variance; useful for multicollinearity and high-dimensional data. **Closed-form:** Yes ($\beta = (X^TX + \lambda I)^{-1}X^Ty$); **Cost:** $O(nd^2 + d^3)$
Lasso Regression	$\min_\beta \sum_{i=1}^n (y_i - X_i\beta)^2 + \lambda\|\beta\|_1$	OLS with ℓ_1 penalty on coefficients	Promotes sparsity; performs variable selection and regularization. **Closed-form:** No; solved by coordinate descent or convex optimization; **Cost:** iterative, $O(ndk)$ for k iterations
Elastic Net	$\min_\beta \sum_{i=1}^n (y_i - X_i\beta)^2$ $+\lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2$	Combines ℓ_1 and ℓ_2 penalties	Balances sparsity and shrinkage; effective when predictors are correlated. **Closed-form:** No; solved by coordinate descent or convex optimization; **Cost:** iterative, $O(ndk)$ for k iterations
Least Absolute Deviations (LAD)	$\min_\beta \sum_{i=1}^n \|y_i - X_i\beta\|$	Minimizes the sum of absolute residuals	Robust to outliers; estimates the conditional median. **Closed-form:** No; solved by linear programming or iterative methods; **Cost:** iterative, $O(ndk)$
Huber Regression	$\min_\beta \sum_{i=1}^n L_\delta(y_i - X_i\beta)$ $L_\delta(r) = \begin{cases} \frac{1}{2}r^2 & \|r\| \le \delta \\ \delta(\|r\| - \frac{1}{2}\delta) & \|r\| > \delta \end{cases}$	Hybrid loss: quadratic for small residuals, linear for large	Robust to outliers while retaining efficiency for small errors. **Closed-form:** No; solved by iterative reweighted least squares (IRLS); **Cost:** iterative, $O(ndk)$
Quantile Regression	$\min_\beta \sum_{i=1}^n \rho_\tau(y_i - X_i\beta)$ $\rho_\tau(r) = r(\tau - \mathbb{I}\{r < 0\})$	Estimates conditional quantiles (e.g., median)	Useful for modeling heterogeneous effects and non-normal errors. **Closed-form:** No; solved by linear programming; **Cost:** iterative, $O(ndk)$
Principal Component Regression (PCR)	OLS on principal components of X	Projects predictors onto principal components before regression	Reduces dimensionality and multicollinearity; interpretable in terms of variance explained. **Closed-form:** Yes (after PCA); **Cost:** $O(nd^2 + d^3)$ for PCA and OLS
Partial Least Squares (PLS)	OLS on latent variables maximizing covariance between X and y	Finds components that explain both predictors and response	Useful when predictors are highly collinear and $p > n$. **Closed-form:** No; solved by iterative algorithms (NIPALS, SIMPLS); **Cost:** iterative, $O(ndk)$
LOWESS/LOESS	$\hat{y}_i = \sum_{j=1}^n w_{ij}y_j$ w_{ij}: local weights from kernel	Locally weighted scatterplot smoothing; fits simple models to local neighborhoods	Captures nonlinear trends without a global parametric form. **Closed-form:** No; local weighted least squares at each point; **Cost:** $O(n^2)$ for n points
Kernel Regression (Nadaraya-Watson)	$\hat{y}(x) = \frac{\sum_{i=1}^n K(x,x_i)y_i}{\sum_{i=1}^n K(x,x_i)}$	Weighted average using a kernel function centered at x	Flexible, smooths data without assuming a parametric form. **Closed-form:** No; direct computation for each x; **Cost:** $O(n)$ per prediction

Continued on next page

Table continued from previous page

Name	Formula	Definition	Significance
Spline Regression	$\min_{f \in \mathcal{S}} \sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int (f''(x))^2 dx$	Fits piecewise polynomials joined at knots; penalizes roughness	Models complex nonlinear relationships with smoothness control. **Closed-form:** Yes (for smoothing splines); **Cost:** $O(n)$ to $O(n^3)$ depending on method
k-Nearest Neighbors Regression	$\hat{y}(x) = \frac{1}{k} \sum_{i \in N_k(x)} y_i$	Averages the k nearest neighbors' responses for prediction	Non-parametric, adapts to local data structure. **Closed-form:** No; requires distance computation for each query; **Cost:** $O(n)$ per prediction