Key Concepts of Kernel Ridge Regression (KRR)

Kernel Ridge Regression (KRR)

Kernel Ridge Regression combines the power of:

- **Ridge Regression** (L2-regularized linear regression)
- **The Kernel Trick** to handle non-linear relationships

1. Ridge Regression

Ridge regression minimizes the sum of squared residuals with an added penalty term for regularization:

$$J(\theta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^{m} \theta_j^2$$

where α is the regularization parameter that controls the strength of regularization.

2. Kernel Trick

The kernel trick allows the model to fit non-linear relationships by implicitly mapping the input data x into a higher-dimensional space using a kernel function $K(x_i, x_j)$.

In Kernel Ridge Regression, we solve the following equation:

$$\hat{y}(x) = \sum_{i=1}^{n} \alpha_i K(x, x_i)$$

where:

- $K(x, x_i)$ is the kernel function that computes the similarity between points x and x_i.
- α_i are the learned coefficients.

Common Kernel Functions

1. **Linear Kernel**:
$$K(x_i, x_j) = x_i^T x_j$$
Equivalent to standard ridge regression.

2. **Polynomial Kernel**:
$$K(x_i, x_j) = (x_i^T x_j + c)^d$$
Allows the model to capture polynomial relationships between features.

3. **Radial Basis Function (RBF) Kernel**:
$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$
Captures complex non-linear relationships and is widely used.

Key Hyperparameters

- **Regularization parameter (α)**: Controls the trade-off between fitting the training data and keeping the model simple.
- **Kernel (K)**: Defines how data is transformed (e.g., 'linear', 'poly', 'rbf').
- **Gamma (γ)**: Parameter for the RBF kernel; controls the smoothness of the decision boundary.
- **Degree (d)**: Degree of the polynomial for polynomial kernels.