CS771 Major Assignment 1 and 2

MLR-48 : Gradient Gang

1 Part - 1

We start with the semi-parametric regression model

$$f(x, z) = p^\top \phi(z)\, x + b,$$

where the kernel associated with ϕ is the polynomial kernel

$$K(z_1, z_2) = (z_1^\top z_2 + c)^d.$$

Feature Map Construction

To convert the model into a purely kernel-based regression, we define an augmented feature map

$$\psi(x, z) = \begin{bmatrix} x\,\phi(z) \\ 1 \end{bmatrix}.$$

Let

$$\tilde{p} = \begin{bmatrix} p \\ b \end{bmatrix}.$$

Then,

$$\tilde{p}^\top \psi(x, z) = p^\top (x\,\phi(z)) + b = p^\top \phi(z)\, x + b,$$

so the model is now linear in the new feature space.

Derived Kernel

The kernel induced by ψ is

$$\tilde{K}((x_1, z_1), (x_2, z_2)) = \langle \psi(x_1, z_1), \psi(x_2, z_2) \rangle.$$

Computing the inner product,

$$\begin{aligned} \tilde{K}((x_1, z_1), (x_2, z_2)) &= \begin{bmatrix} x_1\phi(z_1) \\ 1 \end{bmatrix}^\top \begin{bmatrix} x_2\phi(z_2) \\ 1 \end{bmatrix} \\ &= (x_1\phi(z_1))^\top (x_2\phi(z_2)) + 1 \\ &= x_1 x_2\, \langle \phi(z_1), \phi(z_2) \rangle + 1. \end{aligned}$$

Since

$$\langle \phi(z_1), \phi(z_2) \rangle = K(z_1, z_2),$$

the new kernel becomes

$$\boxed{\tilde{K}((x_1, z_1), (x_2, z_2)) = x_1 x_2\, (z_1^\top z_2 + c)^d + 1.}$$

2 Part - 2

We evaluated Kernel Ridge Regression models using the polynomial kernel

$$K(z_1, z_2) = (z_1^\top z_2 + c)^d,$$

where d is the polynomial degree and c is the bias term. Each model was evaluated on the test set using the R^2 score.

Results

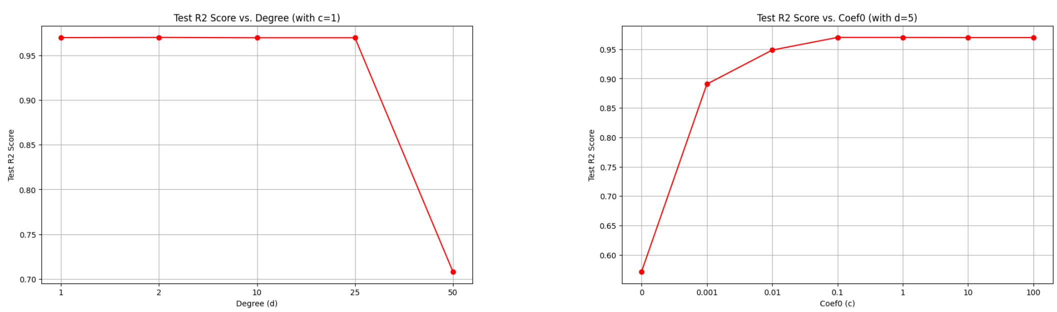
The table below summarizes the combinations of d and c, along with their corresponding test R^2 scores:

Degree d	Test R^2 score
1	0.9696
2	0.9699
5	0.9695
10	0.9695
25	0.9699
50	0.7082
60	-10.5678

(a) Degree d with constant c=1

Coefficient c	Test R^2 score
0	0.5714
0.001	0.8905
0.01	0.9484
0.1	0.9699
1	0.9699
10	0.9696
100	0.9696

(b) Coefficient c with constant d=5



(a) Degree d with constant c=1

(b) Coefficient c with constant d=5

Figure 2: Hyperparameter Analysis

After testing multiple combinations of polynomial kernel parameters, the optimal performance on the provided dataset was obtained with degree d=5 and coefficient c=1. The highest test score was obtained for achieving an R^2 value of 0.9699

Conclusion

From the results, we observe that:

1. Lower degrees d (1 or 2) performed comparably or better than higher-degree polynomials. Very large degrees caused the model to overfit, reducing generalization performance.
2. Small values of c (1 or 2) were sufficient to provide flexibility for the kernel without causing instability.

This suggests that the underlying data does not require a highly complex nonlinear boundary, and lower-order polynomial feature expansions generalize better to unseen data.

3 Part - 4

The XOR Arbiter PUF has two standard Arbiter PUFs. Let the linear models of these two PUFs be $u \in \mathbb{R}^{33}$ and $v \in \mathbb{R}^{33}$.

The model $w \in \mathbb{R}^{1089}$ is the Kronecker product of the individual models given by $w = (\frac{1}{c}.u) \otimes (c.v)$.

For a single Arbiter PUF with 32 stages, the weight vector u is derived linearly from its 128 delays. Let $d_1 \in R^{128}$ be the vector of all delays $(a_i, b_i, c_i, d_i \; for \; i = 0 \ldots 31)$. There exists a sparse matrix $M \in R^{33 \times 128}$ such that:

$$u = Md_1$$

The entries of M are derived from the equations provided $(e.g., w_0 = \alpha_0, ...w_i = \alpha_i + \beta_{i-1})$, where α, β are simple linear combinations of the stage delays.

Similarly, for v we have:

$$v = Md_2$$

To invert these equations, we first need to de-Kroneckerize and recover u and v. If we reshape w into a new 33×33 matrix W, then we have:

$$W = uv^T$$

Now, we can easily recover u and v using Singular Value Decomposition.

$$W = U \Sigma V^T = \sum_i \sigma_i u_i v_i^T$$

We know that W is approximate rank-1. Therefore, we can approximate W as:

$$W \approx \sigma_1 u_1 v_1^T$$

Since $c.u \; and \; (1/c).v$ produce the same product:

$$\hat{u} = \sqrt{\sigma_1} u_1, \hat{v} = \sqrt{\sigma_1} v_1$$

Now, we just need to invert the system. We can set this up as a Ridge Regression optimization problem with constraints.

$$\min_{d_1} ||Md_1 - \hat{u}||_2^2 + \alpha ||d_1||_2^2 \; subject \; to \; d_1 >= 0$$
$$and$$
$$\min_{d_2} ||Md_2 - \hat{v}||_2^2 + \alpha ||d_2||_2^2 \; subject \; to \; d_2 >= 0$$