Formulas

1. **Kernel definition**

$$k(x, x') = \phi(x)^\top \phi(x')$$

2. **Regularized least-squares error (primal)**

$$J(w) = \tfrac{1}{2} \sum_{n=1}^{N} \big(w^\top \phi(x_n) - t_n\big)^2 + \tfrac{\lambda}{2}\, w^\top w$$

3. **Dual representation of** w

$$w = \sum_{n=1}^{N} a_n\, \phi(x_n)$$

4. **Coefficients** a_n

$$a_n = -\tfrac{1}{\lambda}\big(w^\top \phi(x_n) - t_n\big)$$

5. **Primal→dual objective**

$$J(a) = \tfrac{1}{2}\, a^\top \Phi\Phi^\top \Phi\Phi^\top a \;-\; a^\top \Phi\Phi^\top t \;+\; \tfrac{1}{2}\, t^\top t \;+\; \tfrac{\lambda}{2}\, a^\top \Phi\Phi^\top a$$

6. **Gram matrix**

$$K_{nm} = \phi(x_n)^\top \phi(x_m) = k(x_n, x_m)$$

7. **Objective in kernel form**

$$J(a) = \tfrac{1}{2}\, a^\top K K\, a \;-\; a^\top K\, t \;+\; \tfrac{1}{2}\, t^\top t \;+\; \tfrac{\lambda}{2}\, a^\top K\, a$$

8. **Dual solution**

$$a = (K + \lambda I_N)^{-1}\, t$$

9. **Prediction formula**

$$y(x) = k(x)^\top (K + \lambda I_N)^{-1} t\,, \quad k(x)_n = k(x_n, x)$$

10. **Kernel via explicit features**

$$k(x, x') = \sum_{i=1}^{M} \phi_i(x)\, \phi_i(x')$$

11. **Polynomial-square kernel**

$$k(x, z) = (x^\top z)^2$$

12. **Linear kernel special case**

$$\phi(x) = x \implies k(x, x') = x^\top x'$$

13. **Stationary kernel**

$$k(x, x') = k(x - x')$$

14. **Homogeneous (RBF) kernel**

$$k(x, x') = k\big(\|x - x'\|\big)$$

15. **Feature map for** $(x^\top z)^2$

$$\phi(x) = \big(x_1^2,\ \sqrt{2}\, x_1 x_2,\ x_2^2\big)^\top$$

Derivations

Derivation of the Gram Matrix from Regularized Least Squares

Step	Equation	Reason
1	$$J(w) = \tfrac{1}{2}\sum_{n=1}^{N}\big(w^\top \phi(x_n) - t_n\big)^2 \; + \; \tfrac{\lambda}{2}\, w^\top w$$	Primal regularized least-squares objective
2	$$\nabla_w J = \sum_{n=1}^{N}\big(w^\top \phi(x_n) - t_n\big)\, \phi(x_n) + \lambda\, w \; = \; 0$$	Stationarity: set derivative to zero
3	$$w = -\tfrac{1}{\lambda}\sum_{n=1}^{N}\big(w^\top \phi(x_n) - t_n\big)\, \phi(x_n) \; = \; \sum_{n=1}^{N} a_n\, \phi(x_n)$$	Define dual coefficients $a_n = -\frac{1}{\lambda}(w^\top \phi(x_n) - t_n)$
4	$$\Phi = \begin{bmatrix} \phi(x_1)^\top \\ \vdots \\ \phi(x_N)^\top \end{bmatrix}, \quad w = \Phi^\top a$$	Stack feature vectors into design matrix
5	$$\sum_{n=1}^{N}\big(w^\top \phi(x_n) - t_n\big)^2 = w^\top \Big(\sum_{n=1}^{N} \phi(x_n)\phi(x_n)^\top\Big) w - 2\, w^\top \Big(\sum_{n=1}^{N} t_n \phi(x_n)\Big) + \sum_{n=1}^{N} t_n^2$$	Expand $(x-y)^2 = x^2 - 2xy + y^2$
6	$$\sum_{n=1}^{N} \phi(x_n)\phi(x_n)^\top = \Phi^\top \Phi, \quad \sum_{n=1}^{N} t_n \phi(x_n) = \Phi^\top t, \quad \sum_{n=1}^{N} t_n^2 = t^\top t$$	Rewrite sums in matrix form
7	$$J(w) = \tfrac{1}{2}\, w^\top (\Phi^\top \Phi) w - w^\top \Phi^\top t + \tfrac{1}{2}\, t^\top t + \tfrac{\lambda}{2}\, w^\top w$$	Collect terms after substitution and include regularization

Step	Equation	Reason
8	$J(a) = \frac{1}{2}\, a^\top \Phi(\Phi^\top \Phi)\Phi^\top a - a^\top \Phi\Phi^\top t + \frac{1}{2}\, t^\top t + \frac{\lambda}{2}\, a^\top \Phi\Phi^\top a$	Substitute $w = \Phi^\top a$ into $J(w)$
9	$K = \Phi\, \Phi^\top, \quad K_{nm} = \phi(x_n)^\top \phi(x_m)$	Define Gram matrix via matrix multiplication
10	$J(a) = \frac{1}{2}\, a^\top K^2 a - a^\top K\, t + \frac{1}{2}\, t^\top t + \frac{\lambda}{2}\, a^\top K\, a$	Final kernelized objective expressed entirely in terms of K

Kernel symmetry (Formula 1)

Step	Equation	Reason
1	$k(x, x') = \phi(x)^\top \phi(x')$	Definition
2	$k(x', x) = \phi(x')^\top \phi(x)$	Swap arguments
3	$k(x, x') = k(x', x)$	Dot-product symmetry: $a^\top b = b^\top a$

Primal to Dual for regularized LS (Formulas 2–8)

Step	Equation	Reason
1	$J(w) = \frac{1}{2} \sum_n (w^\top \phi_n - t_n)^2 + \frac{\lambda}{2}\, w^\top w$	Given (primal objective)
2	$\nabla_w J = \sum_n (w^\top \phi_n - t_n)\, \phi_n + \lambda\, w = 0$	Differentiate w.r.t. w
3	$\lambda\, w = -\sum_n (w^\top \phi_n - t_n)\, \phi_n$	Rearrange gradient = 0
4	$w = -\frac{1}{\lambda} \sum_n (w^\top \phi_n - t_n)\, \phi_n = \sum_n a_n\, \phi_n$	Define $a_n = -\frac{1}{\lambda}(w^\top \phi_n - t_n)$
5	Substitute $w = \Phi^\top a$ into $J(w) \implies$ $J(a) = \frac{1}{2}\, a^\top \Phi\Phi^\top \Phi\Phi^\top a - a^\top \Phi\Phi^\top t + \frac{1}{2} t^\top t + \frac{\lambda}{2}\, a^\top \Phi\Phi^\top a$	Replace w in objective Algebraic expansion
6	Define $K = \Phi\Phi^\top \implies$ replace in $J(a)$	Gram-matrix substitution
7	$J(a) = \frac{1}{2}\, a^\top K K\, a - a^\top K\, t + \frac{1}{2} t^\top t + \frac{\lambda}{2}\, a^\top K\, a$	Compact kernel form
8	$\nabla_a J = (KK + \lambda K)\, a - K\, t = 0 \implies$ $(K + \lambda I)\, a = t$	Differentiate w.r.t. a, factor K
9	$a = (K + \lambda I_N)^{-1} t$	Solve linear system

Prediction formula (Formula 9)

Step	Equation	Reason
1	$y(x) = w^\top \phi(x)$	Model prediction
2	$w = \Phi^\top a$	From dual representation
3	$y(x) = a^\top \Phi\, \phi(x)$	Substitute w
4	$\Phi\, \phi(x) = k(x)$ with $k(x)_n = k(x_n, x)$	Definition of kernel vector
5	$y(x) = k(x)^\top a = k(x)^\top (K + \lambda I)^{-1} t$	Substitute dual solution for a

---

Kernel via features (Formula 10)

Step	Equation	Reason
1	$\phi(x) \in \mathbb{R}^M$ with $\phi_i(x)$	Assume M basis functions
2	$k(x, x') = \phi(x)^\top \phi(x') = \sum_{i=1}^M \phi_i(x)\, \phi_i(x')$	Dot-product expansion

---

Squared-dot-product kernel (Formulas 11 & 15)

Step	Equation	Reason
1	$k(x, z) = (x^\top z)^2$	Given polynomial kernel
2	Expand: $(x_1 z_1 + x_2 z_2)^2 = x_1^2 z_1^2 + 2 x_1 x_2 z_1 z_2 + x_2^2 z_2^2$	Algebraic square
3	Choose $\phi(x) = (x_1^2,\ \sqrt{2}\, x_1 x_2,\ x_2^2)^\top$	So that $\phi(x)^\top \phi(z)$ matches above

---

Special kernels (Formulas 12–14)

Step	Equation	Reason
1	$\phi(x) = x$	Identity mapping
2	$k(x, x') = x^\top x'$	Linear kernel
3	$k(x, x') = k(x - x')$	Stationarity definition
4	$k(x, x') = k(\lVert x - x' \rVert)$	Homogeneity (radial basis) definition