Source task and conventions

Prove both coefficient and prediction formulas, including a singular kernel.

Conventional solution and proof

Let $K = UD^2U^\top$ on its positive eigenspace, with $U \in \mathbb{R}^{N\times r}$, $D \succ 0$, and $R = UD$. Let $X = UDV^\top$ be a compatible thin SVD. The quadratic-penalty projection theorem makes the feature coefficient $w = V\beta$, where β minimizes

$$\sum_i \{\log(1 + e^{r_i^\top \beta}) - y_i r_i^\top \beta\} + \lambda\|\beta\|^2$$

for the 0/1 logistic convention. A free intercept can be appended consistently. The new inner-product vector is $k_0 = Xx_0 = UDV^\top x_0$. Consequently

$$\alpha = UD^{-1}\beta, \qquad k_0^\top \alpha = x_0^\top VDU^\top UD^{-1}\beta = x_0^\top V\beta = x_0^\top w.$$

This proves (a)–(b) and the new-point linear predictor. Applying the sigmoid gives the predicted probability. For a general PSD kernel, R is an empirical feature matrix and the same optimization defines kernel logistic regression. For singular K, retain only positive eigenvalues or use the pseudoinverse on that support. A legitimate extended kernel has $k_0 \in \operatorname{ran} K$: if $a \in \ker K$, positivity of the block Gram matrix on $(ta, 1)$ for every real t forces $a^\top k_0 = 0$. Thus the omitted zero eigenvalues cause no prediction ambiguity.

Structural or analytic derivation

The nonlinear part of the problem remains the logistic loss; the kernel transformation is linear algebra in a Hilbert-space feature span. Positive regularization makes the slope objective strictly convex, while an unpenalized intercept may need both classes present for a finite optimum.