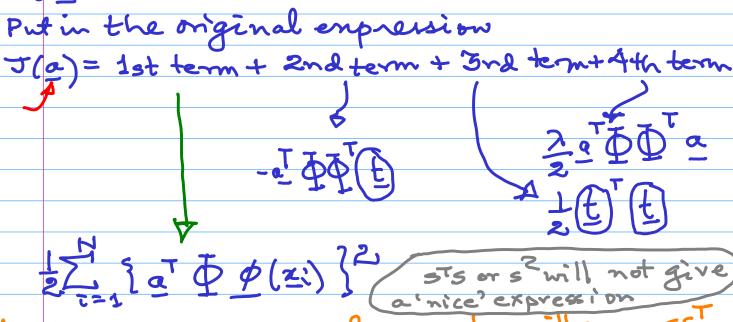
Basic Philosophy

(4) In some cases, formulating an orginal (primal) problem completely in terms of other variables (dual) is possible. There is no guarantee that given a primal problem, it should be possible to formulate a dual one in terms of another

Even if one is able to formulate a dual problem, there is no granewater that the dual problem may have a 'better' solution: better in terms of the computational complexity, attractiveness

in terms of a kernel trick.

Recap:





$$= \frac{1}{2} \sum_{i=1}^{N} \frac{1}{2} \underbrace{a^{T} \Phi \phi(x_{i})}^{2} \underbrace{a^{T} \Phi \phi(x_{i})}^{2$$

Take the fait not involved in the summation,

consider this summation alone

possible ways

D'P

the first term = fa T ΦΦ Φ Φ a

The complete enpression at the optimal value becomes $J(a) = \frac{1}{2} a^{T} (\Phi \Phi^{T}) (\Phi \Phi^{T}) a - a^{T} (\Phi \Phi^{T}) (E)$

we define the gram Motive K & DDT What in this? hot in this:

| Φ^T(±1) | Φ(±1) φ(3z)...φ(zη) |
| Φ^T(±η) | Φ(3z)...φ(zη) |
| Μχη
| Μχη
| Μχη
| Φ^T(±η) | Νχη Now, what is K(i,j)? ith row x jth column $K(i,j) = \phi^{T}(z_{i}) \phi(z_{0})$ $1 \times M \qquad M \times 1$ this is symmetric, Scalar!

= k(ze; zz) the kernel
function K(i,j) = \$ (zi) \$ (zj) = k(zi, zj) Use result: for a quadratic form

o (aTKa) = 2Ka

oa 05(e) - 1.2KKa-K + 2.2Ka = 0 OF KE= K(K+) IN) & assume Ktobe invertible a = (K+ >IN) (t)

4

What is the regression? wt $\phi(z)$ - our model $\gamma(z)$ y(x) = w o(x) For the training data, scalar input we are given target valuesti. y(x.)=wt \phi(x!) is overmodelled out put for which Mother Nature (physical process) has given a value to i.e., for a good model, y(z;)=wTg(zi) should be close to ti. る(2)=ツナダ(2)=(至三)「ダ(と)=三」 (1) consider $\Phi \varphi(z) = \left[\Phi^{\mathsf{T}}(z) \right]$ NYM MX1 \ \phi^{\(\begin{align*}
p(\begin{align*}
p(\begi QT(32) \$(3) p (= 1) \$ (=) $= \left[\begin{array}{c} k\left(\underline{z} + \underline{z}\right) \\ k\left(\underline{z} + \underline{z}\right) \end{array} \right] \longrightarrow \underline{k}(\underline{z})$ $= \left[\begin{array}{c} k\left(\underline{z} + \underline{z}\right) \\ k\left(\underline{z} + \underline{z}\right) \end{array} \right] \longrightarrow \underline{k}(\underline{z})$ $= \left[\begin{array}{c} k\left(\underline{z} + \underline{z}\right) \\ k\left(\underline{z} + \underline{z}\right) \end{array} \right] \longrightarrow \underline{k}(\underline{z})$ => 3(=)= k[(z)(K+ >IN) [E]

Ohysical Dignificance:

(*) The dual formulation allows us to enpress the solution entirely in terms of the kernel function

(*) We recover the original formulation for w; the solution for a can be enpressed as a linear combination of the elements of p (z)

(4) The prediction at z is a linear combination of the target values from the training set.

(4) complexity of the primal and dual formulations

formulations

primal: w = Da typically involve

MXI MXN MXI inverting an MXM

MXI matrix, and

typically, NXXM

Dual: $a = (K + \lambda I_N)^{-1}(t)$ -inverting an $N \times N$ matrix

- complexity-wise, not wise However; the dual formulation in entirely enpressible in terms of the kernel function k (.,.)



If we use the hernelt rck (if it is forsible),
we can work directly with kernels, and avoid
the emphisit introduction of the feature
transformation of (x). This allows us to use
features of high (even infinite)
dimensionality:

CONSTRUCTING KERNEL FUNCTIONS DIRECTLY

Example:
$$k(x, \pm) = (x^T \pm)^2$$

Let $(x, \pm) = (x^T \pm)^2$

Let $(x, \pm) = (x^T \pm)^2$

Let $(x, \pm) = (x, \pm)$

Let $($

try to separate into $\phi(z)$ $\phi(z)$ $\phi(z)$ 2-1