

1

input space: the set of all strings with characters from Σ (Σ^*)
feature space: the frequency of substrings of length p occurring ($\mathbb{R}^{|\Sigma|^p}$)
mapping: ϕ^p

$$K_p(s,t) = \sum_{u \in \Sigma^p} occ(s,u) \cdot occ(t,u) = \langle \phi^p(s), \phi^p(t) \rangle$$

2

dimension feature space: $|\Sigma|^p$

3

ϕ^2	s	t	s,t
AA	0	0	0
AT	2	2	4
AG	0	0	0
AC	0	0	0
TA	1	0	0
TT	0	0	0
TG	2	3	6
TC	0	0	0
GA	0	1	0
GT	0	0	0
GG	0	0	0
GC	2	2	4
CA	1	1	1
CT	1	1	1
CG	0	1	0
CC	0	0	0
K_2			16