- Standard Perceptron:
  - set $w_i=0$ for each feature i
  - set $a^i=0$ for each example i
  - For t=1..T, i=1..n:
    - $y = sign(w \cdot \phi(x^i))$
    - if y ≠ y^i
      - $w = w + y^i \phi(x^i)$
      - $a^i \mathrel{+}= y^i$
  - At all times during learning:
$$w = \sum_k a^k \phi(x^k)$$

- Kernelized Perceptron:
  - set $a^i=0$ for each example i
  - For t=1..T, i=1..n:
    - $y = sign((\sum_k a^k \phi(x^k)) \cdot \phi(x^i))$
      $= sign(\sum_k a^k K(x^k, x^i))$
    - if y ≠ y^i
      - $a^i \mathrel{+}= y^i$

Exactly the same computations, but can use K(u,v) to avoid enumerating the features!!!