# Problem 3

**25 points** In a kernelized linear regression context, prove by **induction** that via gradient descent

At every step t.

Here, *{xi}n* is our training-set examples and each training example *xi ∈ Rd*.

*αi ∈ R* is a Lagrangian multiplier that allows us to express the weights *w ∈ Rd* as a linear combination of the training-set points.

Hints:

1. You want to prove the statement holds for the base cases where *t* = 0 and *t* = 1.
2. Assume that for an update step *t* = *k*, this statement holds.
3. Prove that the statement holds for *t* = *k* + 1.

**Submission** Write your solution in a typeset PDF document (Word or Overleaf/Latex) and name it

**A7\_problem3.pdf** . Hand-writing will not be accepted.

* 1. Base case t = 0;
     1. When t = 0 no update has been done so
        1. Where all for all i, the statement holds for t = 0
  2. Base case t = 1;
     1. When t = 1 the first step is made, the update rule is defined as

Where S is the learning rate and L(w) is the loss function

* + 1. The update rule with , with
    2. Holds for step t=1

1. Inductive step
   1. We assume for an update step t = k, the statement holds
2. Prove that is true for t = k+1
   1. Updated gradient descent rule is;
   2. Substitute where
   3. This is the same form as the rule
      1. Thus the statement holds for all t by induction