**Compressed John Duchi description from optimization perpsective.**

<http://cs229.stanford.edu/extra-notes/boosting.pdf> written by J.C.Duchi.

*“….We now talk about a procedure, known as boosting, which was originally discovered by Rob Schapire, and further developed by* ***Rob Schapire*** *and* ***Yoav Freund****, that automatically chooses feature representations. We take an optimization-based perspective, which is somewhat different from the original interpretation and justification of Freund and Schapire…”*

**Definitions**

where

is a vector of all weak learners.

– is ernomnous dot-product

– is class of the problem

– is called ***margin*** in classification problems. Give knowledge how we confident in it.

m – number of train samples in dataset

**Approach:**

We solve classification problem: via coordinate descend.

T – number of weak learners constructed so far

– edge of weak learner, as bigger value it’s better.

**Algorithm:** Iterate via increasing variable T and make all steps 1-6:

1. **Define weights**
2. **Normalize**
3. **Construct Weak Learner**  to construct predictor with weights

i.e.

*Todo it weak learner should support some analytic algorithm todo it.* ***In fact this is a “coordinate”.***

1. **Compute**
2. **Compute**
3. It can be shown that one dim. Empirical loss minimization via coordinate descend:

Lead to:

**Theorem:**

If (informally it means that at least it’s better then random guessing) then

Boosted classifier have misclassification error less then in all iterations after

**(Notes from J.Duchi, page 6)**

Why bossting is powerfull:

1. It has magic-step-3 which can be considered as that algorithm is adapts to the data and during runtime it’s selected weak learners in automatic way