まず、

コスト関数は、

微分は、

なんで、普通にsquared errorをコスト関数として使わないかというと、squared errorだとnon convexになるかららしい。まず、上の微分の二階微分は、

よって、convexなので、gradient descentで最適化できる。Squared errorの方はまだチェックしてない（実際に計算してチェックしてみよう）。

しかし、**連続値の場合**でも、このコスト関数で良いのか、怪しいので要チェックだ。つまり、の時に、上記のコストが最小値となることを証明できれば良いはず。

**追記：**  
証明できた。

二階微分が非負より、傾きは常に増加するということ。つまり、下に凸なカーブだ。そして、一階微分より、の時に最小値となることが分かる。よって、このコスト関数は、ちゃんとに近づけるという目的を果たすことができる。