

1. You can get better generalization with a stochastic solver
[\[https://arxiv.org/pdf/1708.05070.pdf\]](https://arxiv.org/pdf/1708.05070.pdf)
2. Fastest solution is often through QR factorization, rather than computing inverse or pseudo inverse. Unlike almost every other algorithm in this respect.
3. The reason scaling can still be important is for the optimizer - even though you technically have a convex model and will get the same solution
4. Linear generalization w/ quality feature engineering is stronger than almost every other form of generalization for unstructured data (trees + networks overfit)
5. Best in terms of not overfitting the data - optimal algorithm in low signal-noise ratio environments
6. Only major supervised algorithm with closed form solution
7. Every relationship between your feature and the label should be as close to linear as possible
8. You can use boxcox transform to automatically get close to linear
9. Convex Loss Surface
10. Lasso is not invariant to rescaling
11. L1 penalty leads to laplace distributed coefficients, L2 penalty leads to gaussian distributed coefficients - Bayesian perspective