

Idea: Adaption of theory to networks with biases

Leander Kurscheidt

July 31, 2018

Abstract

1 Affine Transformations

Since our NN "somewhat" resembles linear functions following from the rule $\sigma(x) = \sigma'(x)x$ (i see this rule as separating the linear from the non-linear part), i am curious to see whether ideas of affine spaces and how to integrate them into linear-spaces translate.

I think they do.

2 Idea

Affine subspaces are usually define via $A = v + U_V$, where U_V is a subspace of V and v is a vector of V (see [\(german\) wikipedia](#)). But you can work in them using [homogenous coordinates](#).

Every affine transformation can be turned into a linear transformation with a constant 1-input using the [Augmented Matrix](#) trick.

For example the intercept in linear models in statistics is often modeled this way.

We can adapt this trick with only one additional constant input using this matrix:

$$W' = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \dots & w_{1n} & bias_1 \\ w_{21} & w_{22} & w_{23} & \dots & w_{2n} & bias_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w_{d1} & w_{d2} & w_{d3} & \dots & w_{dn} & bias_n \\ 0 & \dots & \dots & \dots & 0 & 1 \end{bmatrix} = \begin{bmatrix} W & b \\ 0 & \dots & 0 & 1 \end{bmatrix}$$

the last line is the difference to the usual construction and should be ommitted in the last layer.

This results in: $f_{\theta}(x) = f_{\theta'}((x, 1))$,
where θ' is the parameter adpated in the above schema.

Usually proofs get a lot simple when you just have to worry about keeping one input constant. I think a lot of the proofs from the paper should still hold, for example Lemma 2.1 doesn't change.

For this to work, we need to add the additional requirement that $\sigma(1) = 1$

References