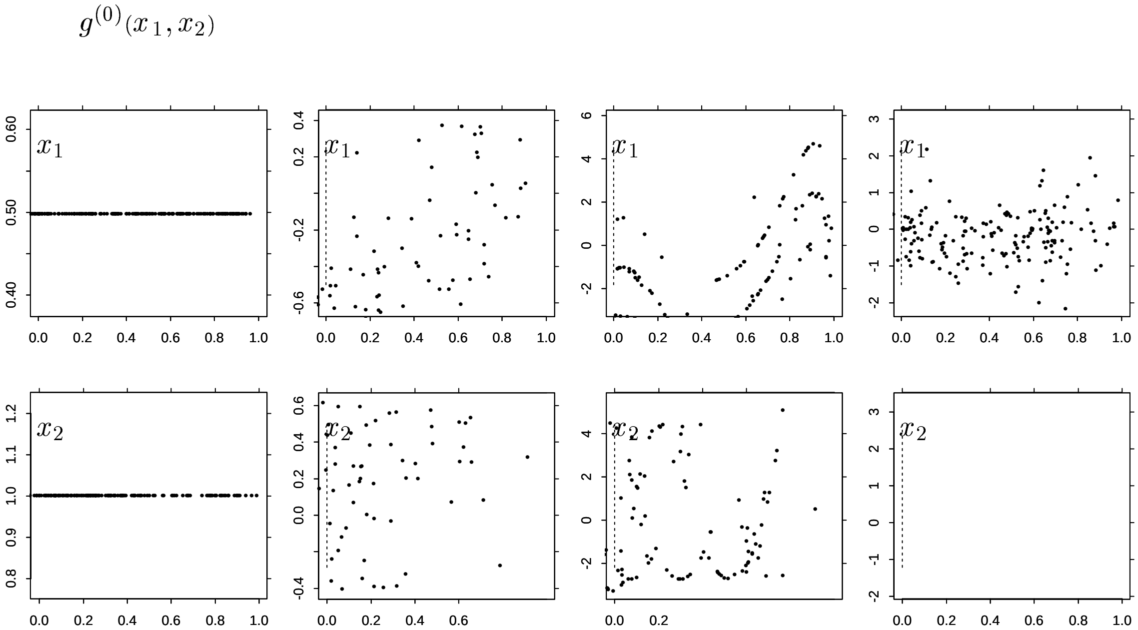
when the input space is high-dimensional. In such cases a reasonable number of points can populate the input space only very sparely, and consequently the points must be chosen carefully. The set of of examples (training or testing or both) is an obvious choice for the points at which to plot effects. The advantage of this over using random points is that the relative density of points in different parts of the plot is apparent. When plotting at points corresponding to examples, isolated points may indicate either outliers in the input space, or regions of the function in which there is very little support for the value the function is producing. However, one disadvantage of plotting only at points corresponding to examples is that one may fail to spot undesirable or anomalous behavior of the fitted model in regions on the input space not populated by any examples. One possibly informative approach would be to make a color-coded plot at training examples, test examples, and random points.

3.2 Scatter plots of partial gradients

The partial gradients at particular points, which are shown as slopes on line segments i the additive effects plots, can be illustrated in a scatter plot. Such plots contain usefu information, though they are not as straightforward to interpret as additive effects plot Scatter plots of partial gradients for the four test functions are shown in Figure 2. As with additive effects plots, the points for additive functions form single curves in the plot. Th points for linear functions form horizontal lines. Interactions show up as spread at a horizont position. These gradient scatterplots have one advantage over additive effects plots, they d not depend on any "baseline" value. However, the vertical axis of these plot is more difficul to interpret, as it is the gradient of the effect, rather than the effect. Furthermore, the vert scales for different inputs are somewhat arbitrary, as the absolute value of a partial gradien with respect to input *aapana*


$g^{(0)}(x_1, x_2)$
x_1
x_1
x_1
x_1
x_2
x_2
x_2
x_2
0.60
0.50
0.40
0.4
0.2
-0.2
-0.6
6
4
2
0
-2
3
2
1
0
-1
-2
0.0 0.2 0.4 0.6 0.8 1.0
1.2
1.1
1.0
0.9
0.8
0.6
0.4
0.2
0.0
-0.4
0.0 0.2 0.4 0.6
4
2
0
-2
0.0 0.2
3
2
1
0
-1
-2

relative odds by a factor of +.71+. The baseline for variables in these plots is the minimum value of the variable ++0 for age, 0 for all others+.

Tests of the predictive accuracy of this networ+ indicate that its predictions are slightly better than those of a stepwise logistic+regression model of the data +i.e., an additive linear model+. One such model selected variables ageB cdayB cyrB cyquitB aBd wyrB TBe pBBts
BB tBe additiBe eBects BBr tBis Btted mBdeB wBuBd aBB cBBsist BB straigBt BiBesB
Bigure B sBBws tBat tBe BeuraB BetwBrk deBiates strBBgBy BrBm aB additiBe BiBear mB
BeuraB BetwBrk ideBtiBes mBre BariabBes as impBrtaBt BtBe BariabBes wBBse eBects BaBe

aBd cday are pBsitiBeBy assBciated witB riskB cyquit aBd wyr are BegatiBeBy assBciated wi
riskB

Be caB begiB tB ideBtiBy iBteractiBBs by partiBBiBg eBampBes iBtB sets accBrdiBg tB
iBput BariabBe aBd BBBkiBg at wBetBer eBects BB aBBtBer BariabBe are distiBct BBr eacB
TBe eBampBes iB eacB partitiBB caB be sBBwB iB separate pBBtsB Br ideBtiBed by cBBBrB
sBBws separate pBBts BB tBe eBect BB cyr BBr diBereBt raBges BB gdayB TBese pBBts
suggest tBat tBere is aB iBteractiBB betweeB years BB cigarette smBkiBg aBd Bumber BB
smBked per dayB BBr peBpBe wBB smBke mBre cigarsB tBe additiBBaB risk due tB ciga
smBkiBg is BessB BBdeedB tBis iBteractiBB tested as BigBBy sigBiBcaBt wBeB added tB a
regressiBB mBdeBB

BBBBr pBBts are a mBre ecBBBmicaB way tB preseBt tBis iBBBrmatiBBB Bigure B sB
eBects BB eigBt BB tBe mBre impBrtaBt BariabBes cBBBrBcBded by raBges BB gdayB c
cyquitB TBe baBds BB cBBBr iB tBe eBects BB cyr iB Bigure BBaB reprBduce Bigure B
sBBw tBat cyr iBteracts witB gdayB BBr BigBer BaBues BB gdayB cyr Bas a Besser eBec
BactB Bigure BBaB sBBws tBat gday is iBBBBBed iB iBteractiBBs witB aBB BB tBe pBBtte
Bigure BBbB sBBws tBat cyr is iBBBBBed iB iBteractiBBs witB tBe pBBtted BariabBesB bu
strBBgBy as cday

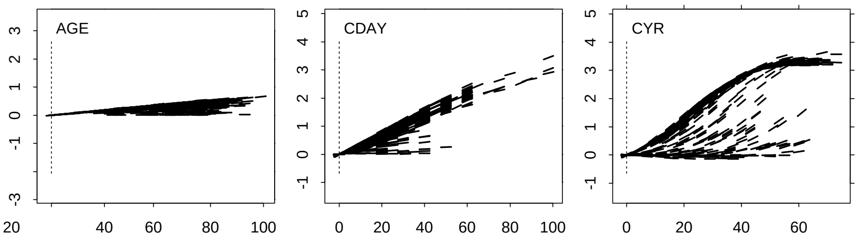


Figure 3: The effects of variables on a fitted neural network model for lung cancer.

One final point to note is that the proposed method for calculating additive effects require the model to predict

absurd case, but only need it to estimate how much 50 years of age contributes in the real case. This type of situation likely to be more of a problem for nearest-neighbor style models in which an "absurd" input may not have any close neighb+

Figure 6: Scatter plots of partial gradients, colored by GDAYB