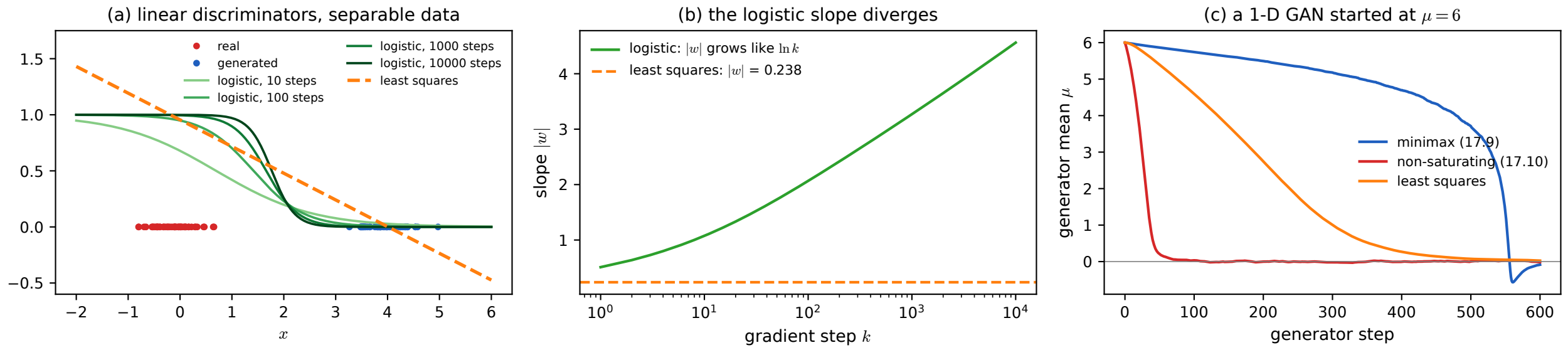

(a) linear discriminators, separable data
real
generated
logistic, 10 steps
logistic, 100 steps
logistic, 1000 steps
logistic, 10000 steps
least squares
1.5
1.0
0.5
0.0
−0.5
−2
−1
0
1
2
3
4
5
6
x
(b) the logistic slope diverges
logistic: $|w|$ grows like $\ln k$
least squares: $|w|$ = 0.238
slope $|w|$
4
3
2
1
10^0
10^1
10^2
10^3
10^4
gradient step k
(c) a 1-D GAN started at $\mu = 6$
generator mean μ
6
5
4
3
2
1
0
minimax (17.9)
non-saturating (17.10)
least squares
0
100
200
300
400
500
600
generator step