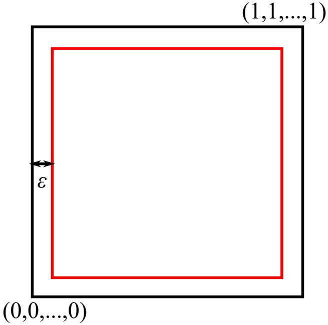
More on Logistic Regression

Foundations of Data Analysis

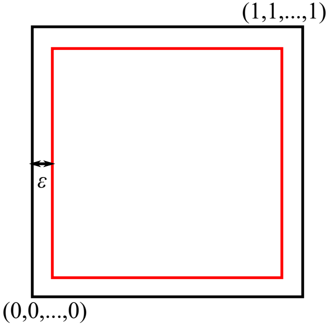
April 7, 2022

Volumes in High Dimensions



What is the volume of the unit d-cube shrunk by some small amount in each dimension?

Volumes in High Dimensions

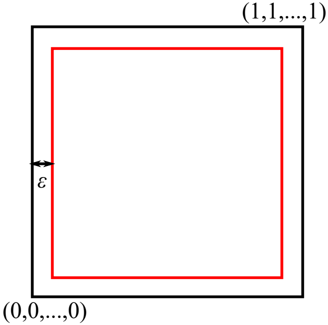


What is the volume of the unit d-cube shrunk by some small amount in each dimension?

$$V = (1 - 2\epsilon)^d$$

Approaches 0 as $d \to \infty$

Volumes in High Dimensions



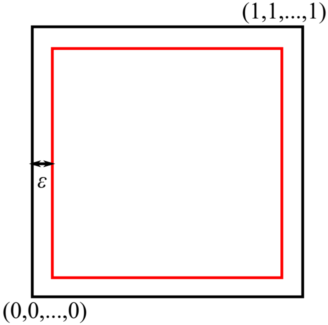
What is the volume of the unit d-cube shrunk by some small amount in each dimension?

$$V = (1 - 2\epsilon)^d$$

Approaches 0 as $d \to \infty$

Example: $256 \times 256 \times 3$ images, $\epsilon = \frac{1}{256}$

Volumes in High Dimensions



What is the volume of the unit d-cube shrunk by some small amount in each dimension?

$$V = (1 - 2\epsilon)^d$$

Approaches 0 as $d \to \infty$

Example: $256 \times 256 \times 3$ images, $\epsilon = \frac{1}{256}$

$$V \approx 2.0 \times 10^{-670}$$

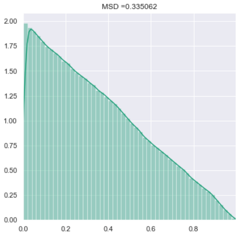
Distances in High Dimensions

Sample two points uniformly from the unit d-cube:
$X, Y \sim \mathrm{Unif}([0,1]^d)$

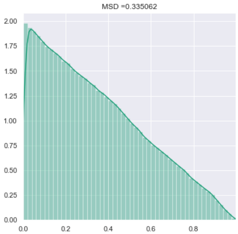
Distances in High Dimensions

Sample two points uniformly from the unit d-cube:
$X, Y \sim \mathrm{Unif}([0,1]^d)$

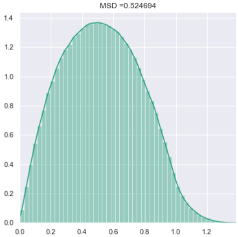
What is the distribution of the distance between them?
$D = \|X - Y\|$



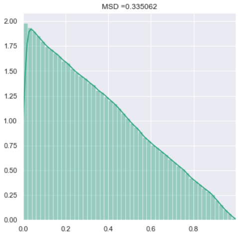
$$d = 1$$


MSD =0.335062
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8

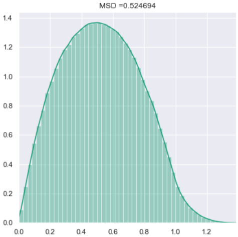
$d = 1$


MSD =0.524694
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1.2

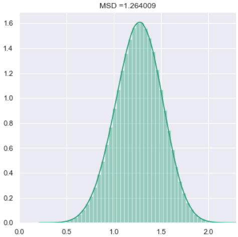
$d = 2$



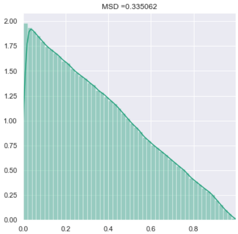
$d = 1$



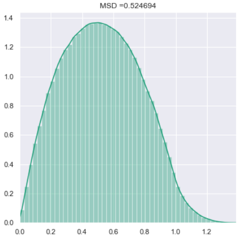
$d = 2$



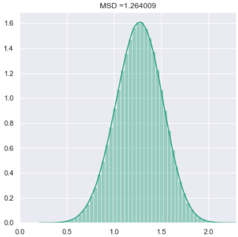
$d = 10$



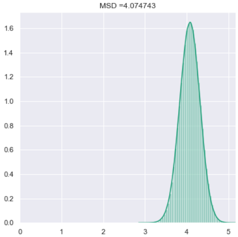
$d = 1$



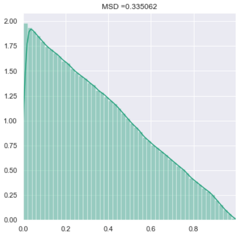
$d = 2$



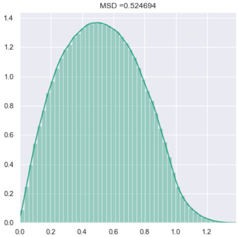
$d = 10$



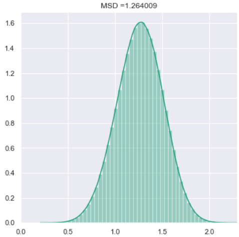
$d = 100$



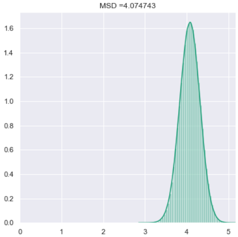
$d = 1$



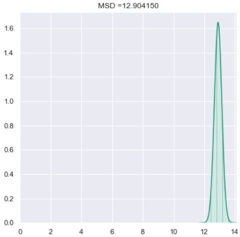
$d = 2$



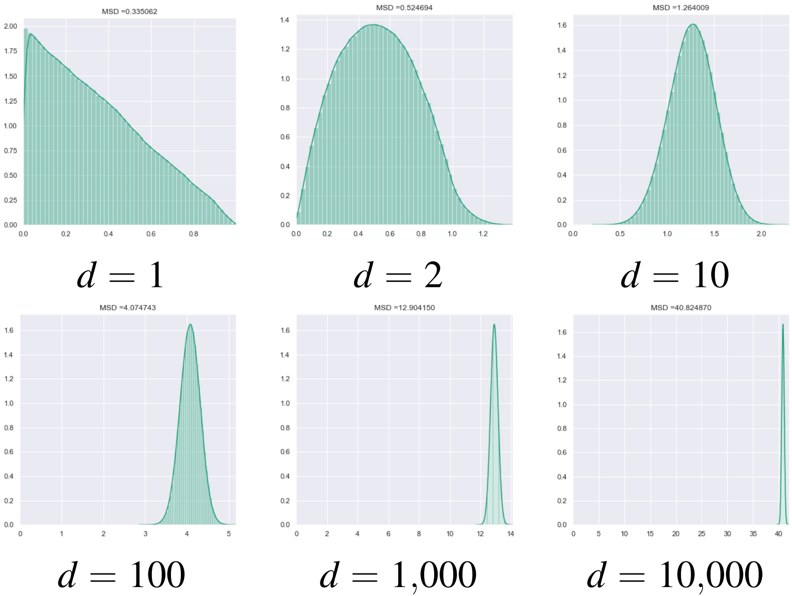
$d = 10$



$d = 100$



$d = 1{,}000$


MSD =0.335062
MSD =0.524694
MSD =1.264009
d = 1
d = 2
d = 10
MSD =4.074743
MSD =12.904150
MSD =40.824870
d = 100
d = 1,000
d = 10,000

Angles in High Dimensions

Sample two directions uniformly from the unit d-sphere:
$X, Y \sim \mathrm{Unif}(S^d)$

Angles in High Dimensions

Sample two directions uniformly from the unit d-sphere:
$X, Y \sim \mathrm{Unif}(S^d)$

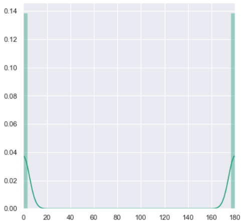
What is the distribution of the angle between them?
$A = \arccos\langle X, Y \rangle$

Angles in High Dimensions

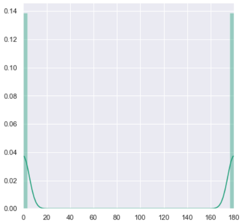
Sample two directions uniformly from the unit d-sphere:
$X, Y \sim \mathrm{Unif}(S^d)$

What is the distribution of the angle between them?
$A = \arccos\langle X, Y \rangle$

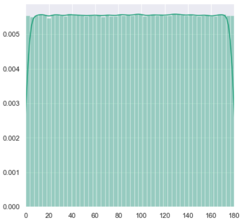
Note: Equivalently, sample $X, Y \sim N(0, I)$ and normalize: $A = \arccos\left\langle \frac{X}{\|X\|}, \frac{Y}{\|Y\|} \right\rangle$



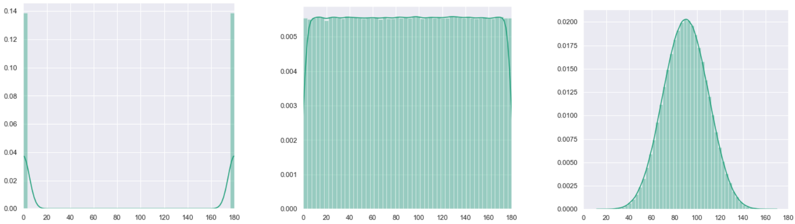
$d = 1$


0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
20
40
60
80
100
120
140
160
180

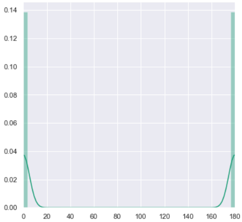
$d = 1$


0.005
0.004
0.003
0.002
0.001
0.000
0
20
40
60
80
100
120
140
160
180

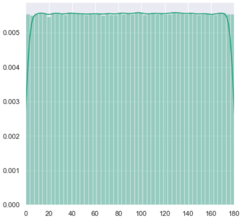
$d = 2$



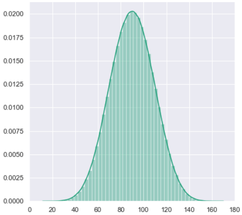
$d = 1$ $\quad$ $d = 2$ $\quad$ $d = 10$



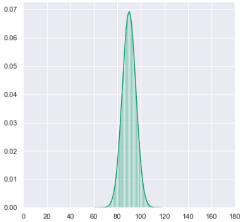
$d = 1$



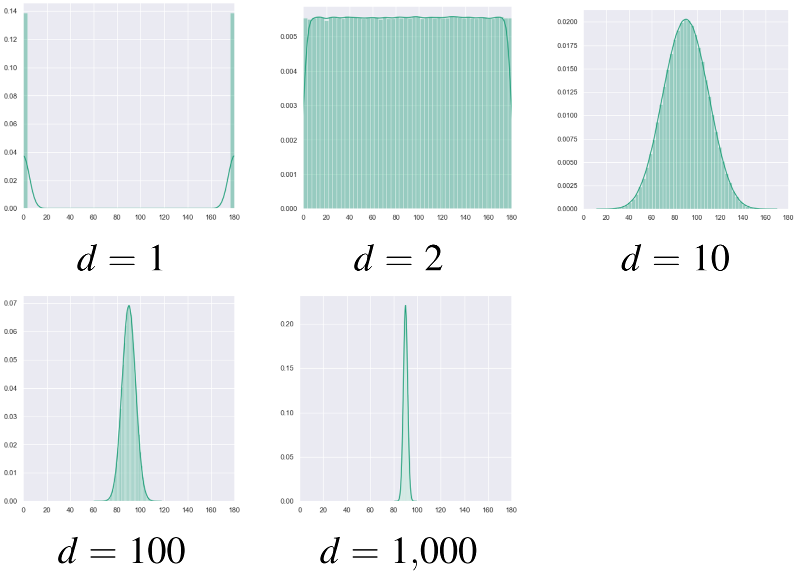
$d = 2$



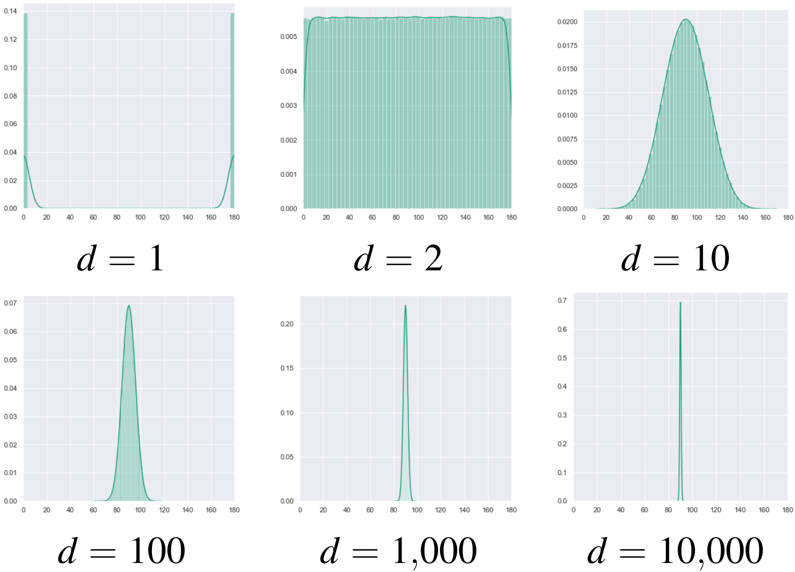
$d = 10$



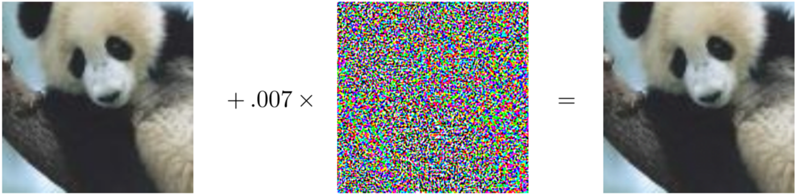
$d = 100$



d = 1
d = 2
d = 10
d = 100
d = 1,000


$d = 1$
$d = 2$
$d = 10$
$d = 100$
$d = 1{,}000$
$d = 10{,}000$

Adversarial Examples



$\boldsymbol{x}$
"panda"
57.7% confidence

$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"nematode"
8.2% confidence

$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"gibbon"
99.3 % confidence

Goodfellow et al. ICLR 2015

High-Dimensionality Explanation?

The Relationship Between High-Dimensional Geometry and Adversarial Examples

Justin Gilmer, Luke Metz, Fartash Faghri, Samuel S. Schoenholz, Maithra Raghu, Martin Wattenberg, & Ian Goodfellow
Google Brain
{gilmer,lmetz,schsam,maithra,wattenberg,goodfellow}@google.com
faghri@cs.toronto.edu

High-Dimensionality Explanation?

The Relationship Between High-Dimensional Geometry and Adversarial Examples

Justin Gilmer, Luke Metz, Fartash Faghri, Samuel S. Schoenholz, Maithra Raghu, Martin Wattenberg, & Ian Goodfellow
Google Brain
{gilmer,lmetz,schsam,maithra,wattenberg,goodfellow}@google.com
faghri@cs.toronto.edu

ARE ADVERSARIAL EXAMPLES INEVITABLE?

Ali Shafahi, Ronny Huang, Christoph Studer, Soheil Feizi & Tom Goldstein

High-Dimensionality Explanation?

The Relationship Between High-Dimensional Geometry and Adversarial Examples

Justin Gilmer, Luke Metz, Fartash Faghri, Samuel S. Schoenholz, Maithra Raghu, Martin Wattenberg, & Ian Goodfellow
Google Brain
{gilmer,lmetz,schsam,maithra,wattenberg,goodfellow}@google.com
faghri@cs.toronto.edu

ARE ADVERSARIAL EXAMPLES INEVITABLE?

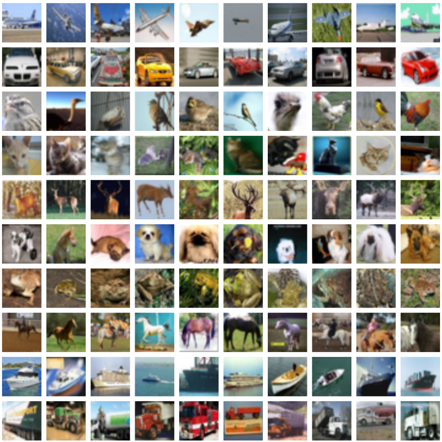
Ali Shafahi, Ronny Huang, Christoph Studer, Soheil Feizi & Tom Goldstein

The Curse of Concentration in Robust Learning:
Evasion and Poisoning Attacks from Concentration of Measure

Saeed Mahloujifar* Dimitrios I. Diochnos† Mohammad Mahmoody‡

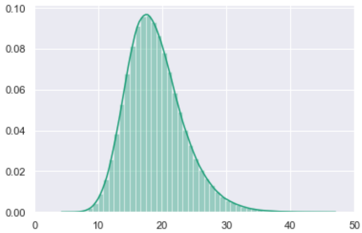
CIFAR-10

airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck

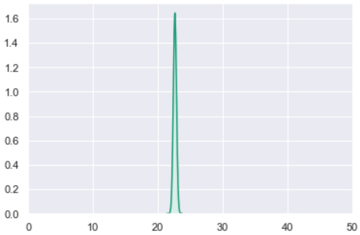


$32 \times 32 \times 3 = 3{,}072$ dimensions
10 classes

Distances in Real Data

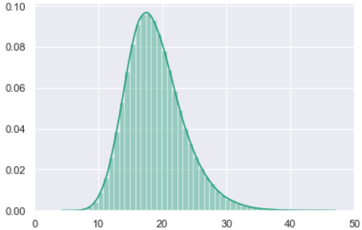


CIFAR-10

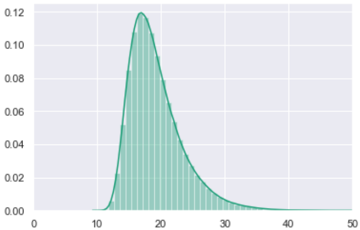


$\mathrm{Unif}([0,1]^{3072})$

Distances in Real Data



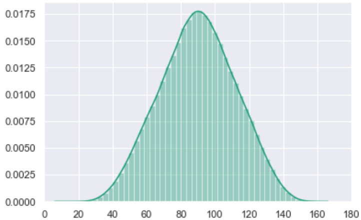
CIFAR-10



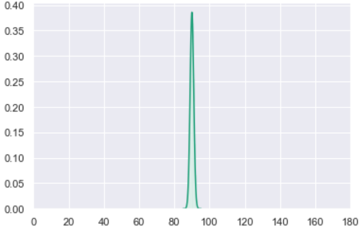
$N(0, S)$

$S =$ sample covariance of CIFAR-10

Angles in Real Data

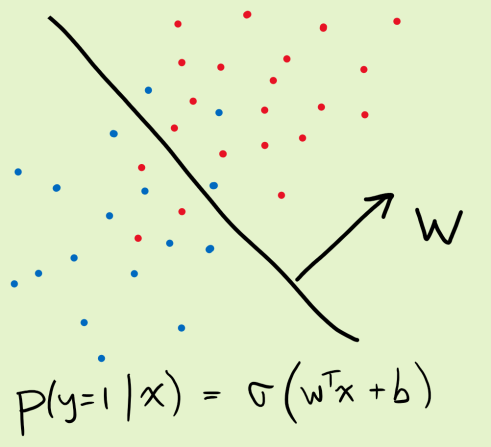


CIFAR-10



$N(0, I)$

Logistic Regression



Planes vs. Frogs: Test Images

Planes vs. Frogs: Test Images

Logistic regression accuracy = 89.40%

Gradient Attack

Move input x in direction that increases loss function, J:

Gradient Attack

Move input x in direction that increases loss function, J:

Attack: $x + \eta$
$\eta = \lambda \nabla_x J(w, x, y), \quad$ for some $\lambda > 0$

Gradient Attack

Move input x in direction that increases loss function, J:

Attack: $x + \eta$
$\eta = \lambda \nabla_x J(w, x, y), \quad$ for some $\lambda > 0$

For logistic regression: $\eta \propto w$

Planes vs. Frogs

Test Images Projected onto w

89.40% Accuracy

Planes vs. Frogs

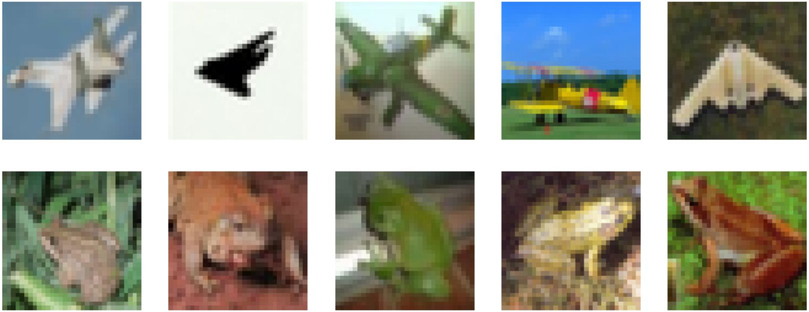
Test Images Projected onto w

89.40% Accuracy

Gradient attack of $1.5\frac{w}{\|w\|}$

50.25% Accuracy

Planes vs. Frogs: Test Images



Accuracy = 89.40%

Planes vs. Frogs: Gradient Attack

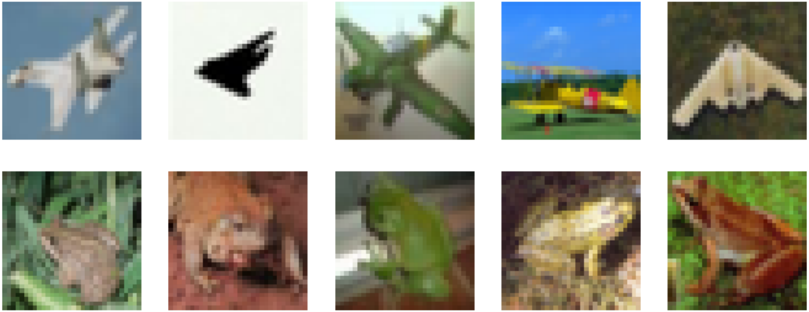
Accuracy = 50.25%

Random Attack

Add a random vector η to an image x

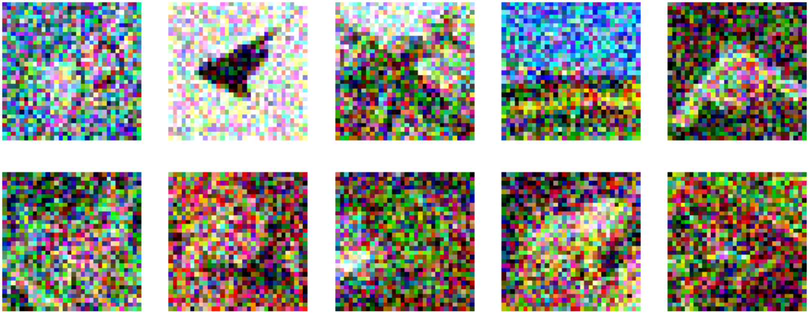
$$\eta \sim \mathrm{Unif}(-0.5, 0.5)^{3072}$$

Planes vs. Frogs: Test Images

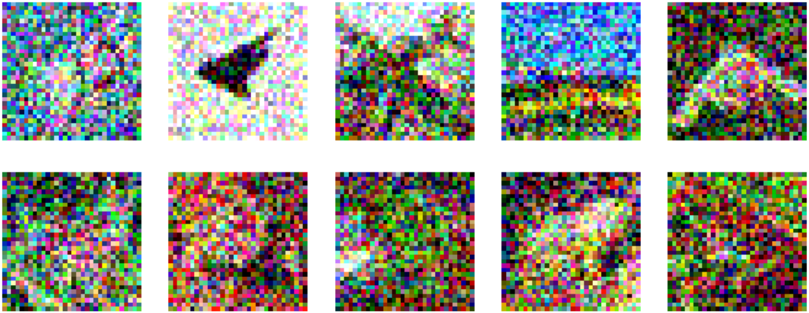


Accuracy = 89.40%

Planes vs. Frogs: Noise

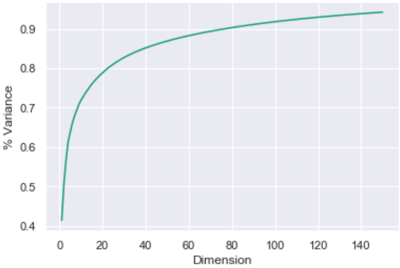


Planes vs. Frogs: Noise



Accuracy = 88.50%

How Many Dimensions Do We Need?



1. Compute PCA of training data
2. Project onto top 10 dimensions
3. Retrain logistic regression

Planes vs. Frogs: Test Images

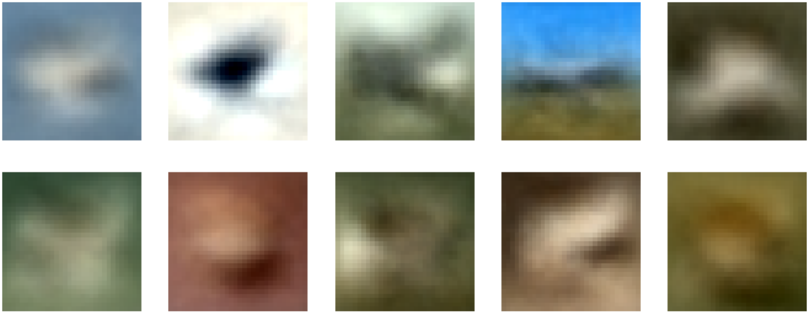
Accuracy = 89.40%

Planes vs. Frogs: PCA $(d = 10)$

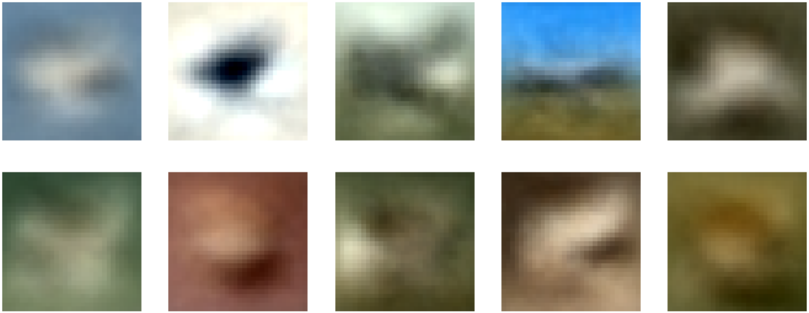
Planes vs. Frogs: PCA $(d = 10)$

Accuracy = 86.75%

Planes vs. Frogs: PCA $(d = 20)$

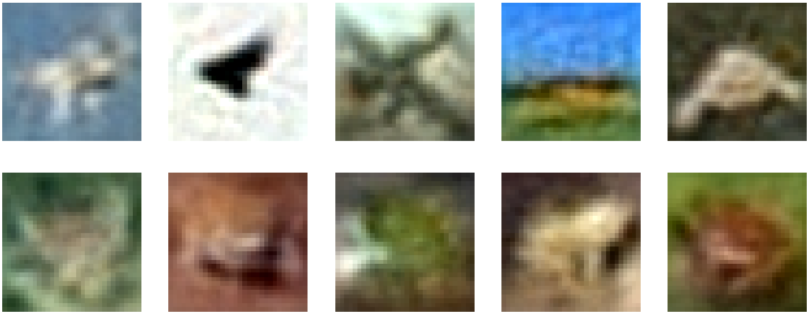


Planes vs. Frogs: PCA $(d = 20)$

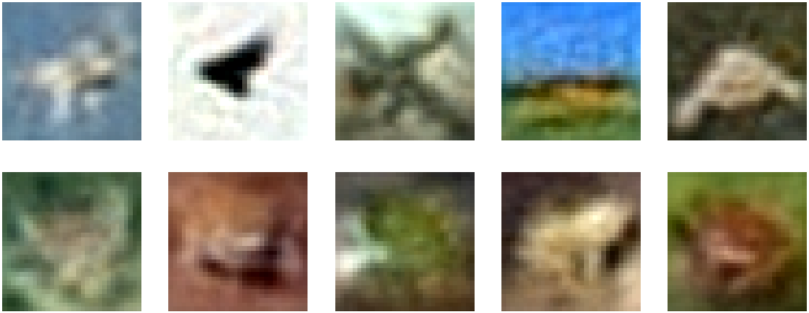


Accuracy = 88.60%

Planes vs. Frogs: PCA $(d = 100)$

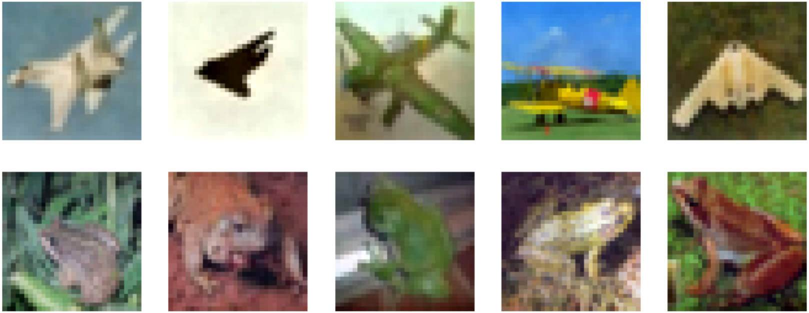


Planes vs. Frogs: PCA $(d = 100)$

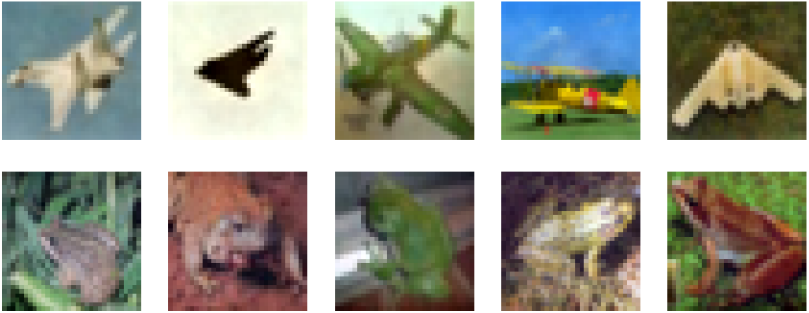


Accuracy = 89.30%

Gradient Attack: PCA $(d = 10)$

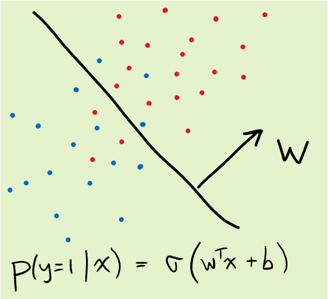


Gradient Attack: PCA $(d = 10)$



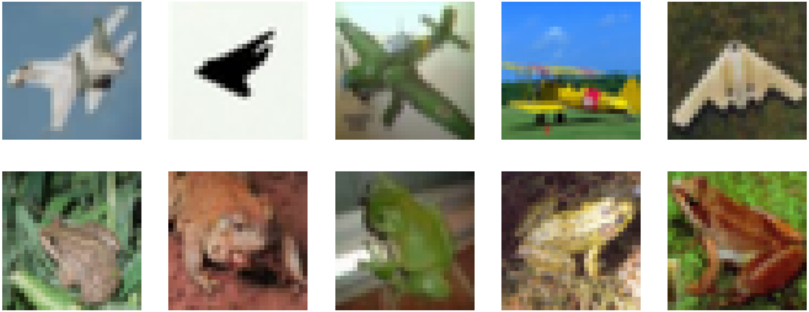
Accuracy = 50.80 %, **but** $||\eta|| = 2.4$ (vs. 1.5 before)

Removing The “Best” Separating Dimension



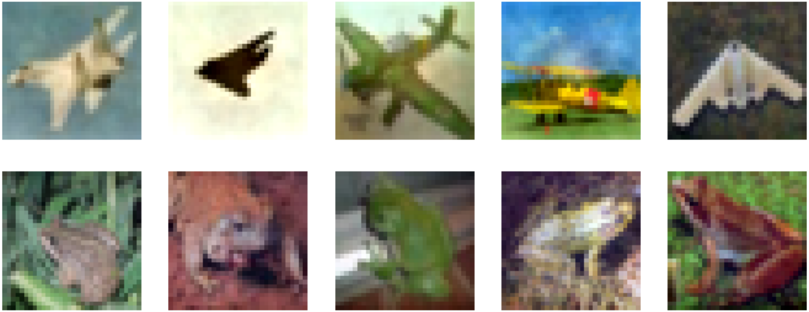
1. Project out the w dimension found by logistic regression
2. Retrain logistic regression
3. Run on original test data (without the projection step)

Planes vs. Frogs: Test Images



Accuracy = 89.40%

Planes vs. Frogs: Remove w



Accuracy = 86.00%

Manifold Hypothesis

Real data lie near lower-dimensional manifolds