

Adversarial Machine Learning

A comprehensive overview

Abhinav Venkataraman

Samsung SDS Research America

Table of contents

1. Introduction
2. Titleformats
3. Elements
4. Conclusion

Introduction

What's an Adversarial Example?

Machine learning models that misclassify examples that are slightly different (sometimes even imperceptible from human eye) from correctly classified examples drawn from the data distribution.

Problem Definition

Regular Neural Network Training

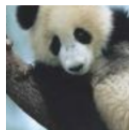
Train a model on a dataset such that you take the gradient of loss function w.r.t model parameters. In this way, you maximize on the score of the correct class.

Adversarial Learning

Generate an image by doing the following.

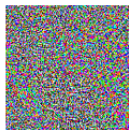
- Taking gradient of loss function w.r.t input image of the class you want to perturb.
- Take your original image and perturb it by a small quantity η along the direction of gradient.
- Run the model along this result and see the result.

Adversarial Training Example



x
"panda"
57.7% confidence

+ .007 ×



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

=



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

Titleformats

metropolis supports 4 different titleformats:

- Regular
- SMALLCAPS
- ALLSMALLCAPS
- ALLCAPS

They can either be set at once for every title type or individually.

This frame uses the `smallcaps` titleformat.

Potential Problems

Be aware, that not every font supports small caps. If for example you typeset your presentation with pdfTeX and the Computer Modern Sans Serif font, every text in smallcaps will be typeset with the Computer Modern Serif font instead.

This frame uses the `allsmallcaps` titleformat.

Potential problems

As this titleformat also uses smallcaps you face the same problems as with the `smallcaps` titleformat. Additionally this format can cause some other problems. Please refer to the documentation if you consider using it.

As a rule of thumb: Just use it for plaintext-only titles.

This frame uses the `allcaps` titleformat.

Potential Problems

This titleformat is not as problematic as the `allsmallcaps` format, but basically suffers from the same deficiencies. So please have a look at the documentation if you want to use it.

Elements

The theme provides sensible defaults to
`\emph{emphasize}` text, `\alert{accent}` parts
or show `\textbf{bold}` results.

becomes

The theme provides sensible defaults to *emphasize* text, **accent** parts or
show **bold** results.

Font feature test

- Regular
- *Italic*
- SMALLCAPS
- **Bold**
- **Bold Italic**
- **Bold SmallCaps**
- Monospace
- *Monospace Italic*
- Monospace Bold
- *Monospace Bold Italic*

Items

- Milk
- Eggs
- Potatos

Enumerations

1. First,
2. Second and
3. Last.

Descriptions

PowerPoint Meeh.

Beamer Yeeeha.

- This is important

- This is important
- Now this

- This is important
- Now this
- And now this

- This is really important
- Now this
- And now this

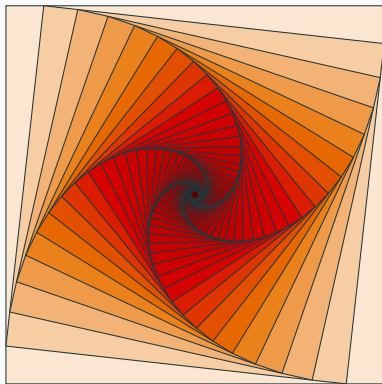


Figure 1: Rotated square from texample.net.

Table 1: Largest cities in the world (source: Wikipedia)

City	Population
Mexico City	20,116,842
Shanghai	19,210,000
Peking	15,796,450
Istanbul	14,160,467

Three different block environments are pre-defined and may be styled with an optional background color.

Default

Block content.

Alert

Block content.

Example

Block content.

Default

Block content.

Alert

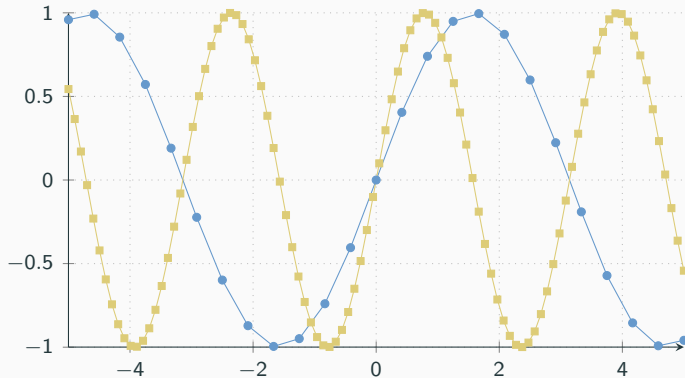
Block content.

Example

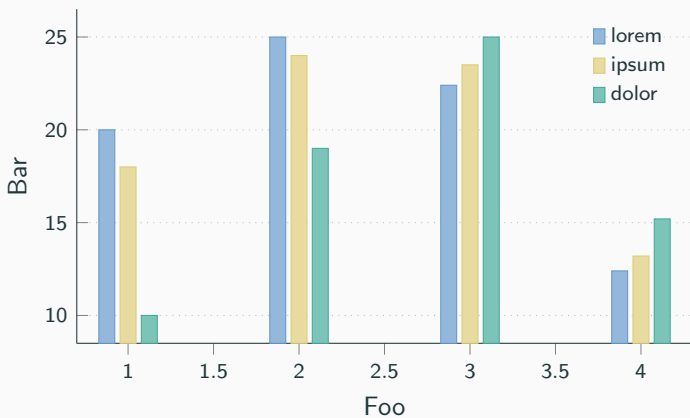
Block content.

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

Line plots



Bar charts



Veni, Vidi, Vici

metropolis defines a custom beamer template to add a text to the footer. It can be set via

```
\setbeamertemplate{frame footer}{My custom footer}
```

Some references to showcase `[allowframebreaks]` [4, 2, 5, 1, 3]

Conclusion

Get the source of this theme and the demo presentation from

`github.com/matze/mtheme`

The theme *itself* is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



Questions?

Backup slides

Sometimes, it is useful to add slides at the end of your presentation to refer to during audience questions.

The best way to do this is to include the `appendixnumberbeamer` package in your preamble and call `\appendix` before your backup slides.

metropolis will automatically turn off slide numbering and progress bars for slides in the appendix.

References I



P. Erdős.

A selection of problems and results in combinatorics.

In *Recent trends in combinatorics (Matrahaza, 1995)*, pages 1–6.
Cambridge Univ. Press, Cambridge, 1995.



R. Graham, D. Knuth, and O. Patashnik.

Concrete mathematics.

Addison-Wesley, Reading, MA, 1989.



G. D. Greenwade.

The Comprehensive Tex Archive Network (CTAN).

TUGBoat, 14(3):342–351, 1993.



D. Knuth.

Two notes on notation.

Amer. Math. Monthly, 99:403–422, 1992.



H. Simpson.

Proof of the Riemann Hypothesis.

preprint (2003), available at

<http://www.math.drofnats.edu/riemann.ps>, 2003.