Predicting the Future using Deep Adversarial Networks

Learning With No Labeled Data

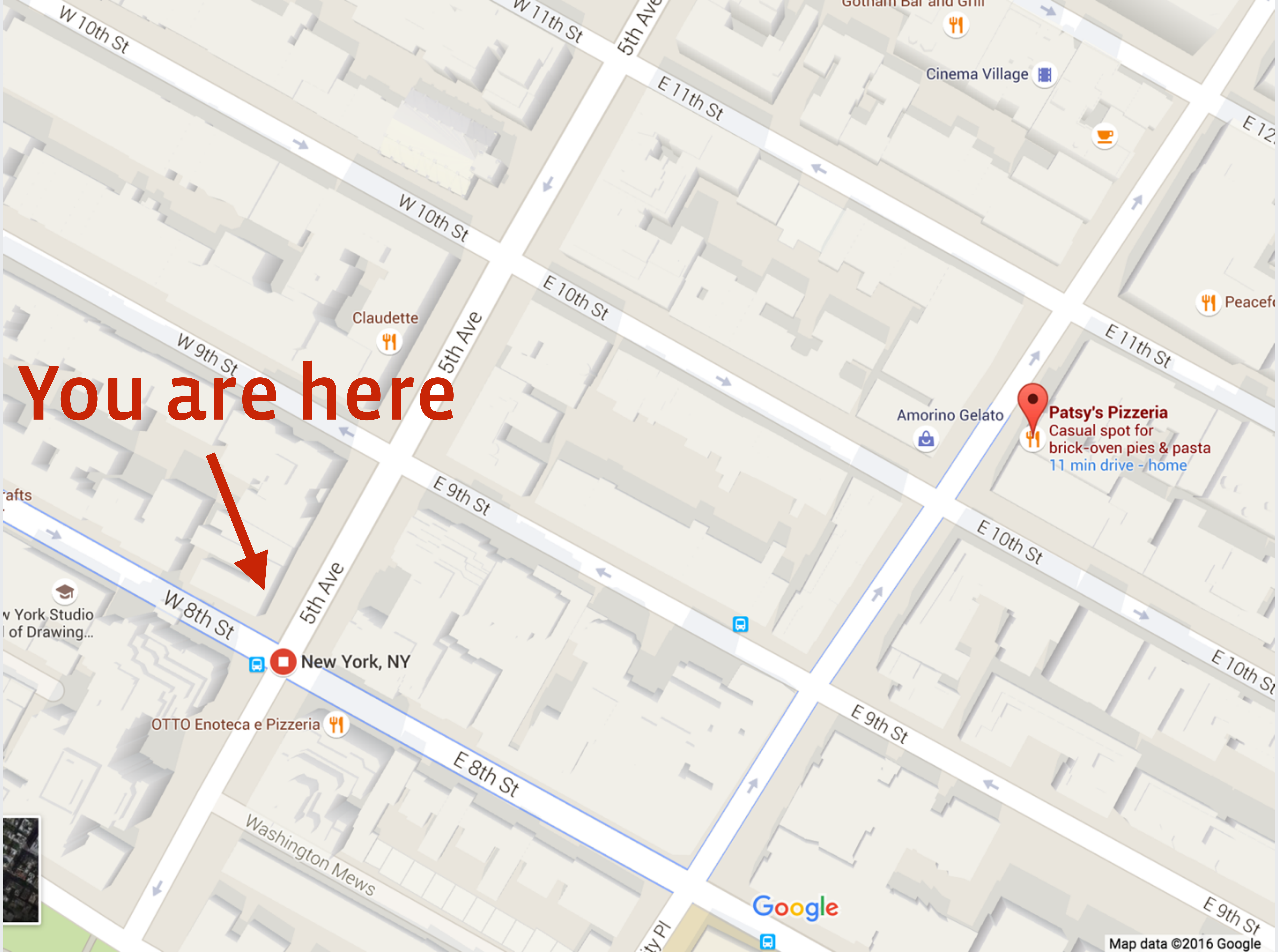
Soumith Chintala

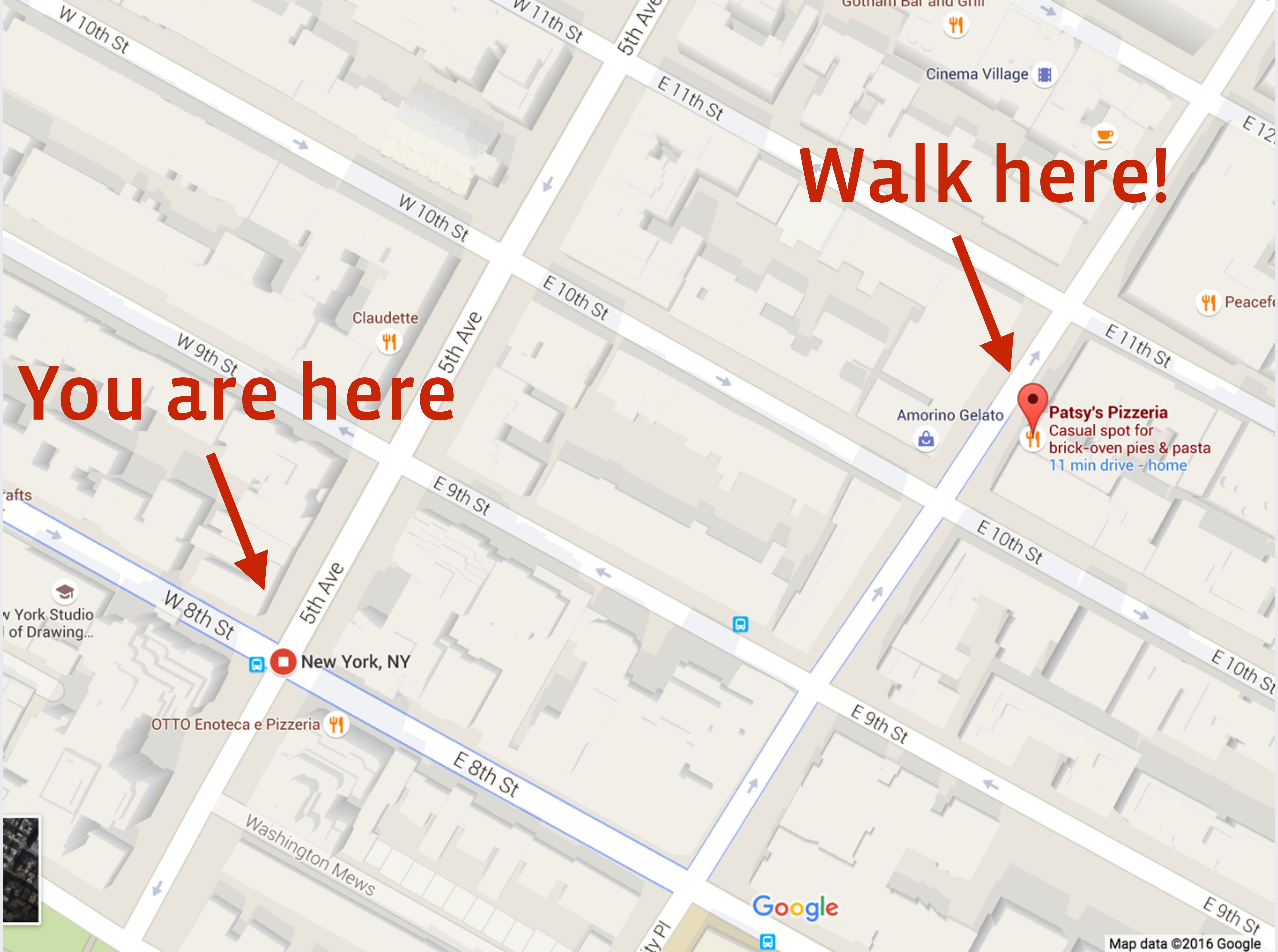
Facebook AI Research

Overview

of the talk

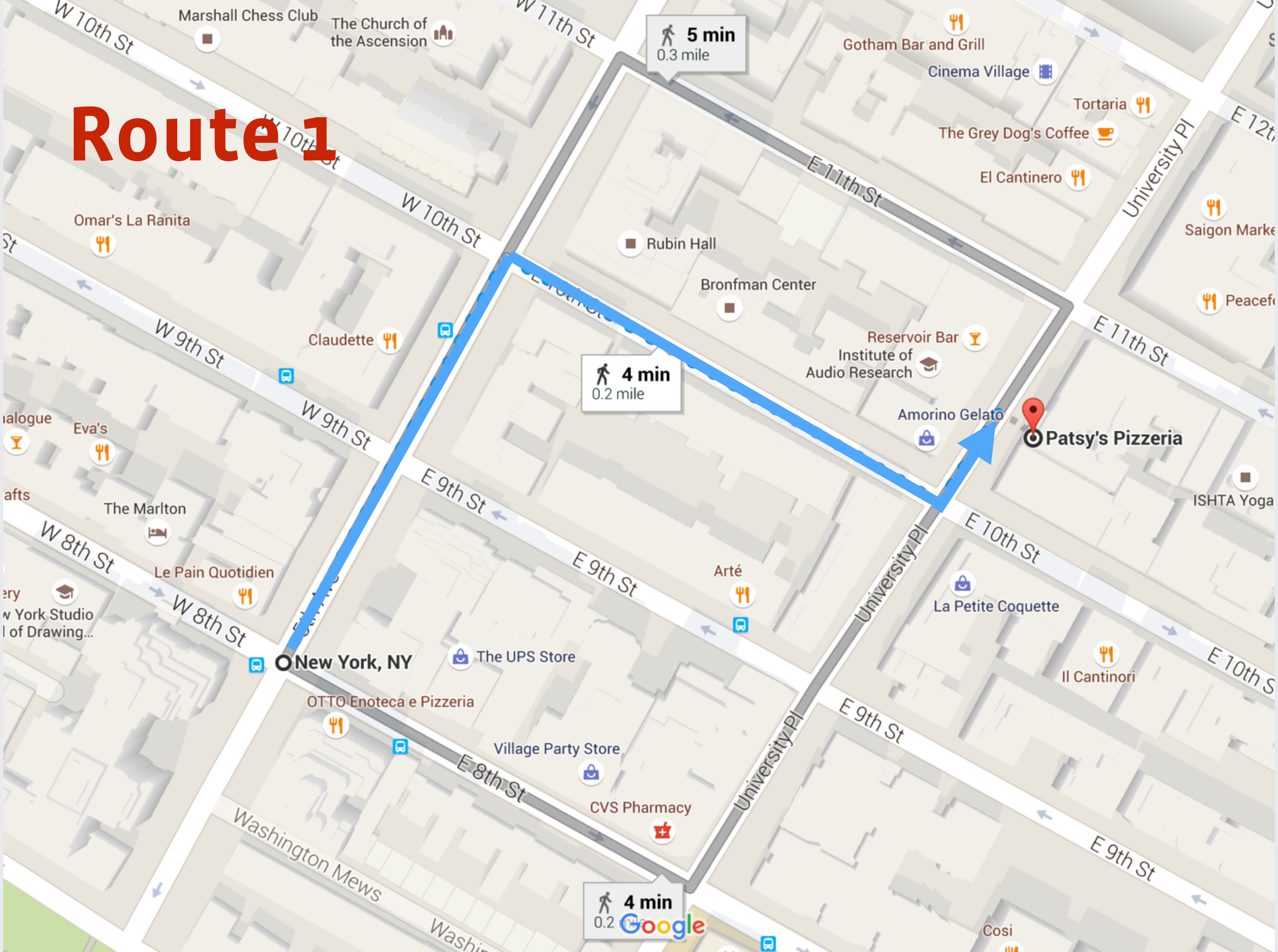
- The problem at hand
- What are the benefits?
- How did we solve it
- What have we achieved
- What’s left?


You are here
W 10th St
W 11th St
5th Ave
Gotham Bar and Grill
Cinema Village
E 11th St
E 12
W 10th St
E 10th St
Claudette
W 9th St
5th Ave
Peacef
E 11th St
Amorino Gelato
Patsy's Pizzeria
Casual spot for
brick-oven pies & pasta
11 min drive - home
E 9th St
afts
E 10th St
5th Ave
W 8th St
w York Studio
of Drawing...
New York, NY
E 10th St
OTTO Enoteca e Pizzeria
E 8th St
E 9th St
Washington Mews
Google
E 9th St
Map data ©2016 Google

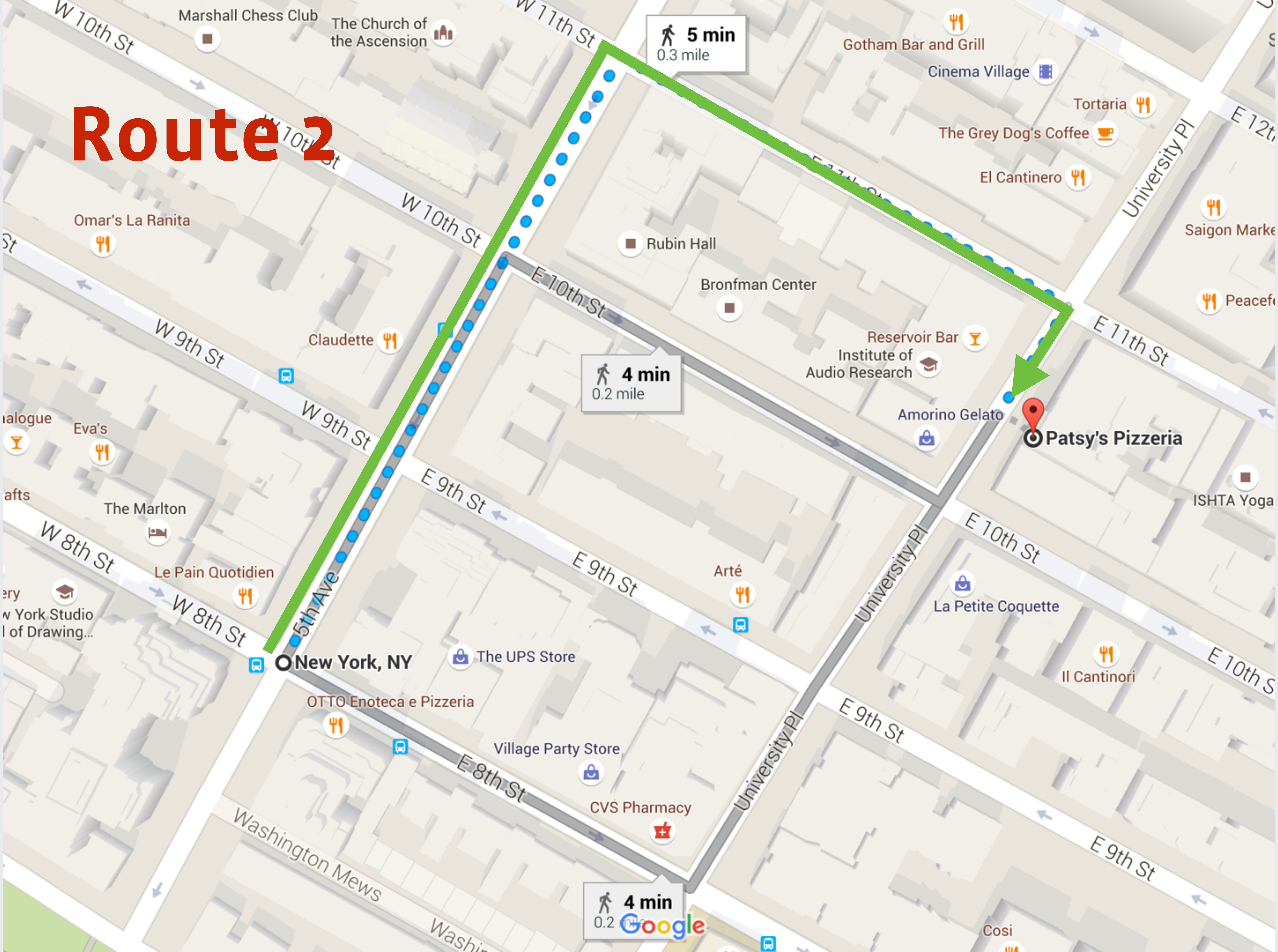


Walk here!
You are here
W 10th St
W 11th St
5th Ave
E 11th St
Cinema Village
E 10th St
W 9th St
Claudette
Amorino Gelato
Patsy's Pizzeria
Casual spot for
brick-oven pies & pasta
11 min drive - home
E 11th St
E 9th St
W 8th St
New York, NY
OTTO Enoteca e Pizzeria
E 8th St
Washington Mews
E 10th St
E 9th St
Google
Map data ©2016 Google

Route 1


5 min
0.3 mile
4 min
0.2 mile
4 min
New York, NY
Patsy's Pizzeria
Marshall Chess Club
The Church of the Ascension
Gotham Bar and Grill
Cinema Village
Tortaria
The Grey Dog's Coffee
El Cantinero
Omar's La Ranita
Rubin Hall
Bronfman Center
Claudette
Reservoir Bar
Institute of Audio Research
Amorino Gelato
Eva's
The Marlton
Le Pain Quotidien
The UPS Store
OTTO Enoteca e Pizzeria
Village Party Store
CVS Pharmacy
Arté
La Petite Coquette
Il Cantinori
ISHTA Yoga
Cosi
W 10th St
W 11th St
E 11th St
W 9th St
E 9th St
W 8th St
E 8th St
E 10th St
University Pl
Washington Mews
Google

Route 2



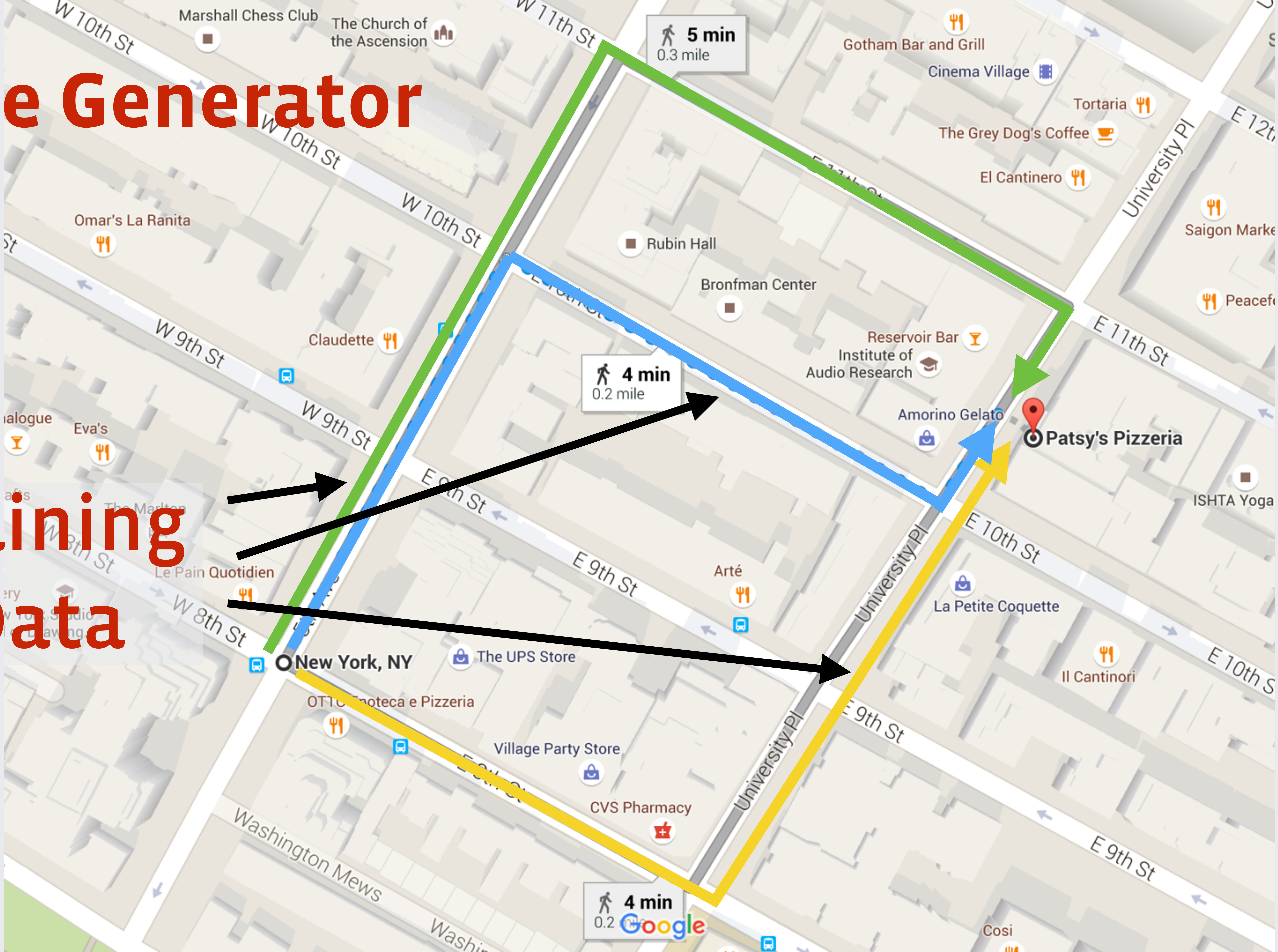
5 min
0.3 mile
4 min
0.2 mile
4 min
0.2
Google
New York, NY
Patsy's Pizzeria
W 10th St
W 11th St
W 9th St
W 8th St
E 10th St
E 11th St
E 9th St
E 8th St
E 12th
5th Ave
University Pl
Washington Mews
Marshall Chess Club
The Church of the Ascension
Gotham Bar and Grill
Cinema Village
Tortaria
The Grey Dog's Coffee
El Cantinero
Saigon Marke
Peacef
Omar's La Ranita
Rubin Hall
Bronfman Center
Reservoir Bar
Institute of Audio Research
Amorino Gelato
ISHTA Yoga
Claudette
Eva's
The Marlton
Le Pain Quotidien
Arté
La Petite Coquette
Il Cantinori
The UPS Store
OTTO Enoteca e Pizzeria
Village Party Store
CVS Pharmacy
Cosi

Route 3

5 min
0.3 mile
4 min
0.2 mile
4 min
0.2 mile
W 10th St
W 11th St
W 9th St
W 8th St
E 11th St
E 10th St
E 9th St
E 8th St
5th Ave
University Pl
Washington Mews
Marshall Chess Club
The Church of the Ascension
Gotham Bar and Grill
Cinema Village
Tortaria
The Grey Dog's Coffee
El Cantinero
Saigon Marke
Omar's La Ranita
Rubin Hall
Bronfman Center
Claudette
Reservoir Bar
Institute of Audio Research
Amorino Gelato
Patsy's Pizzeria
Eva's
The Marlton
Le Pain Quotidien
ISHTA Yoga
Arté
La Petite Coquette
New York, NY
The UPS Store
Il Cantinori
Village Party Store
CVS Pharmacy
Cosi
Google

Let’s Train a Route Generator

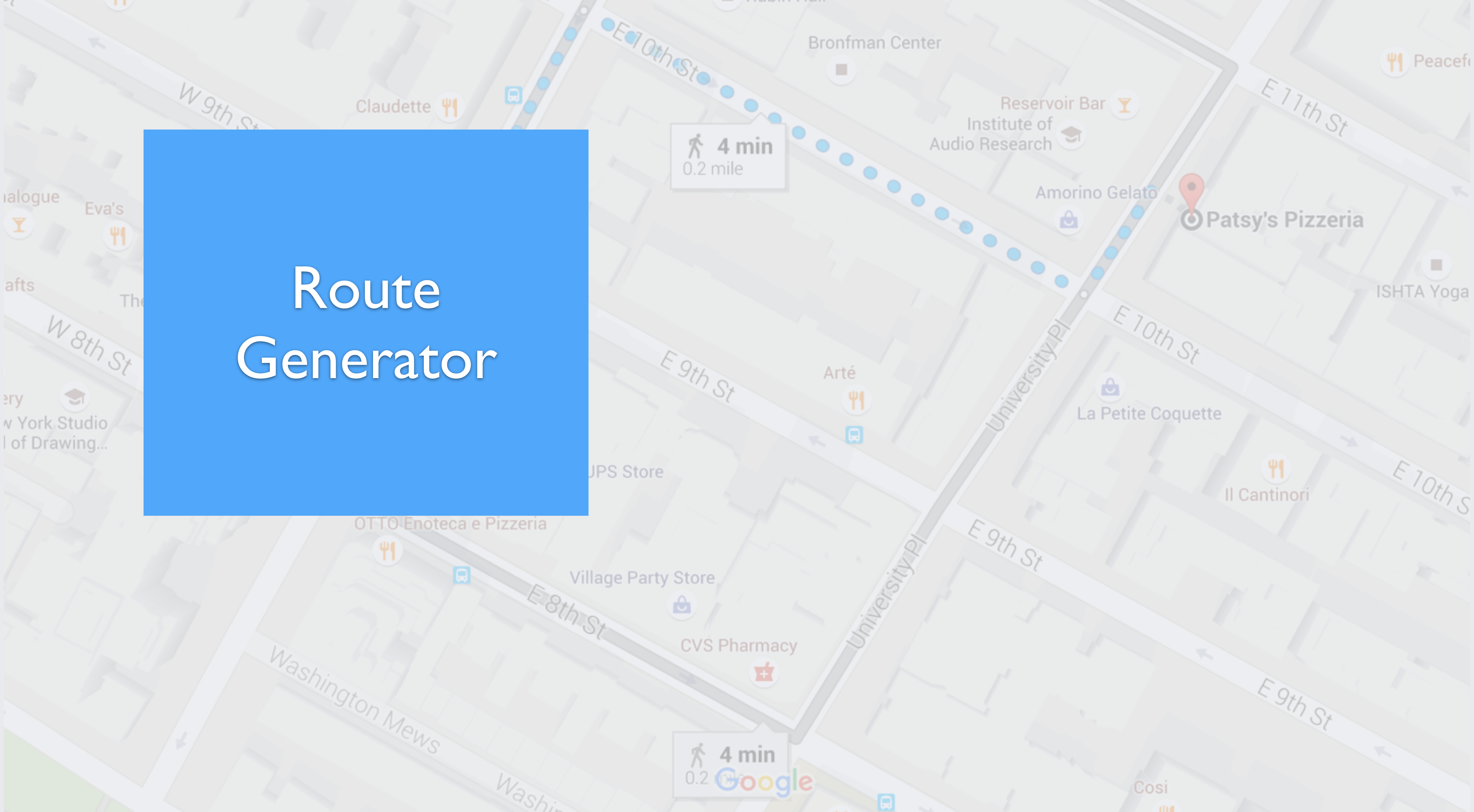
Route Generator



Training Data

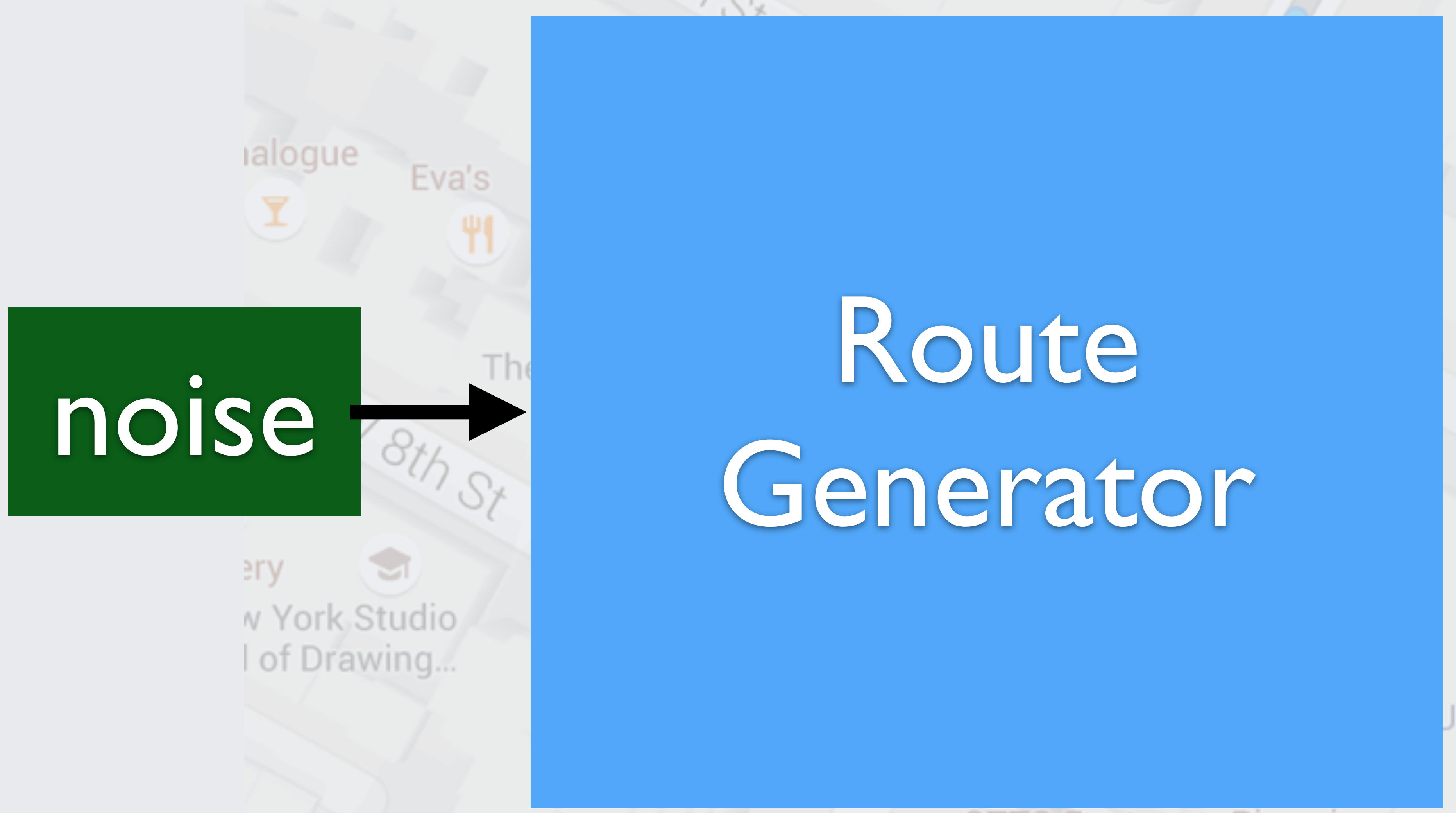
Route Generator

Linear Regressor with Mean-square Error



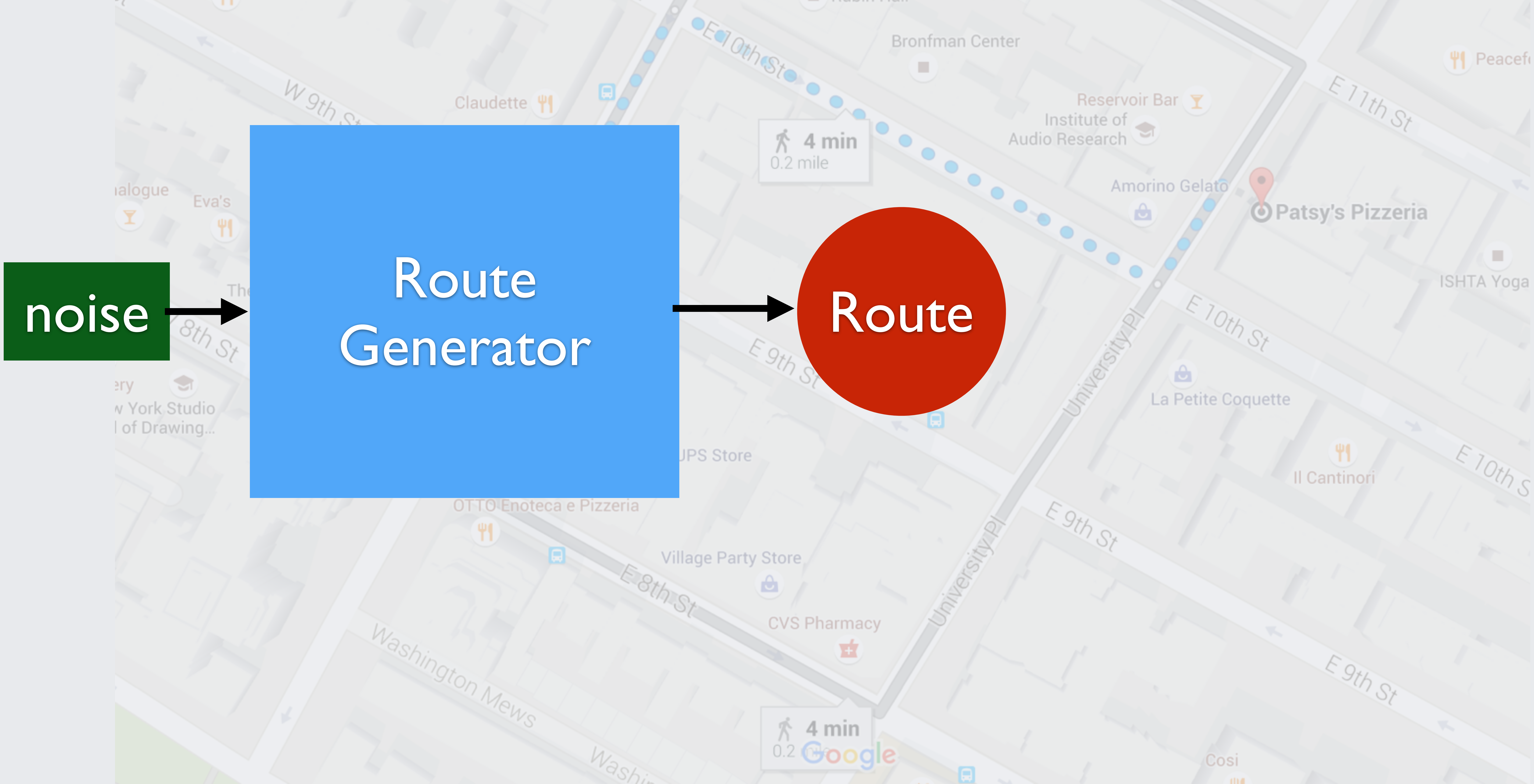
Route Generator

Linear Regressor with Mean-square Error



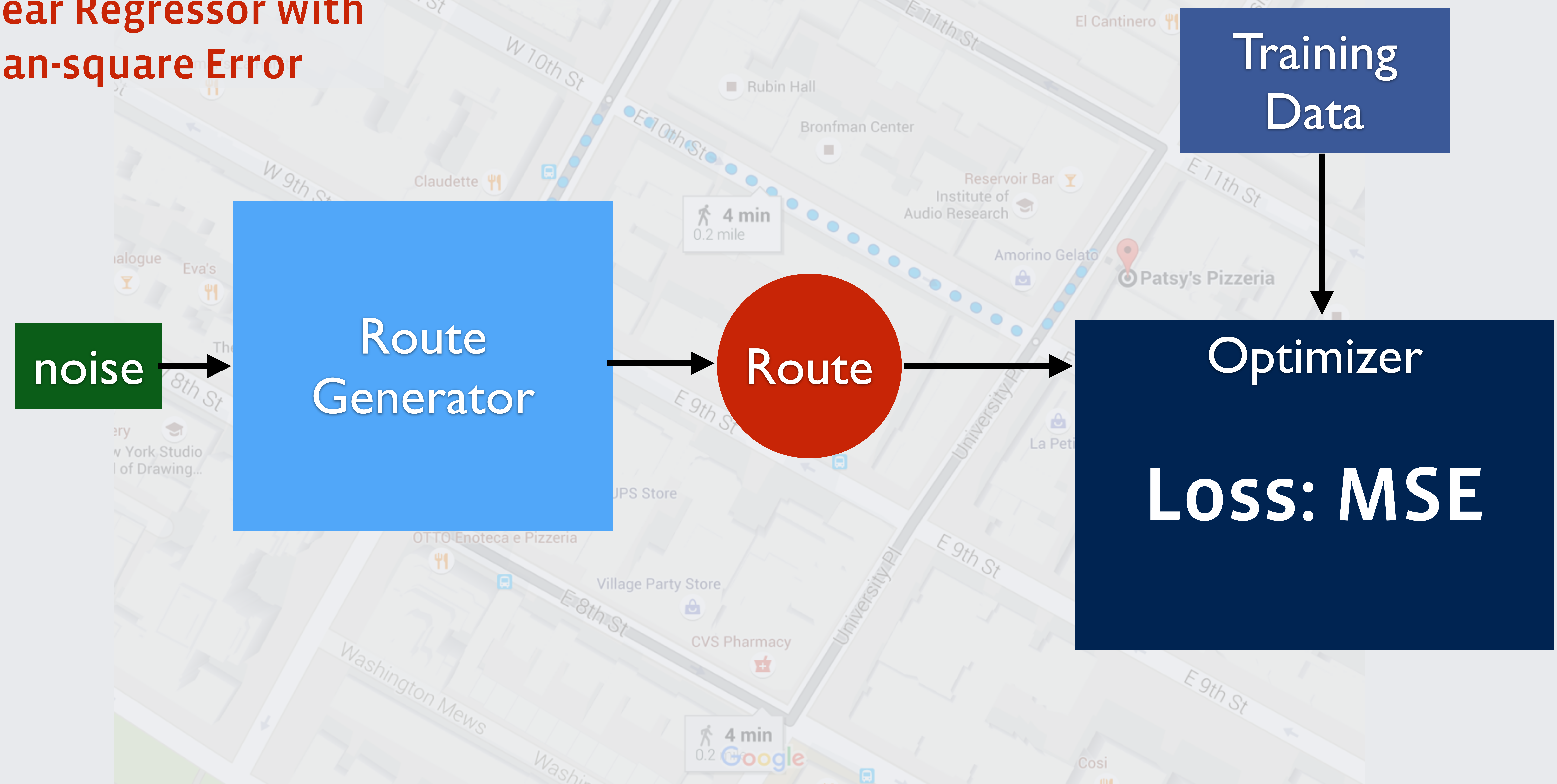
Route Generator

Linear Regressor with Mean-square Error



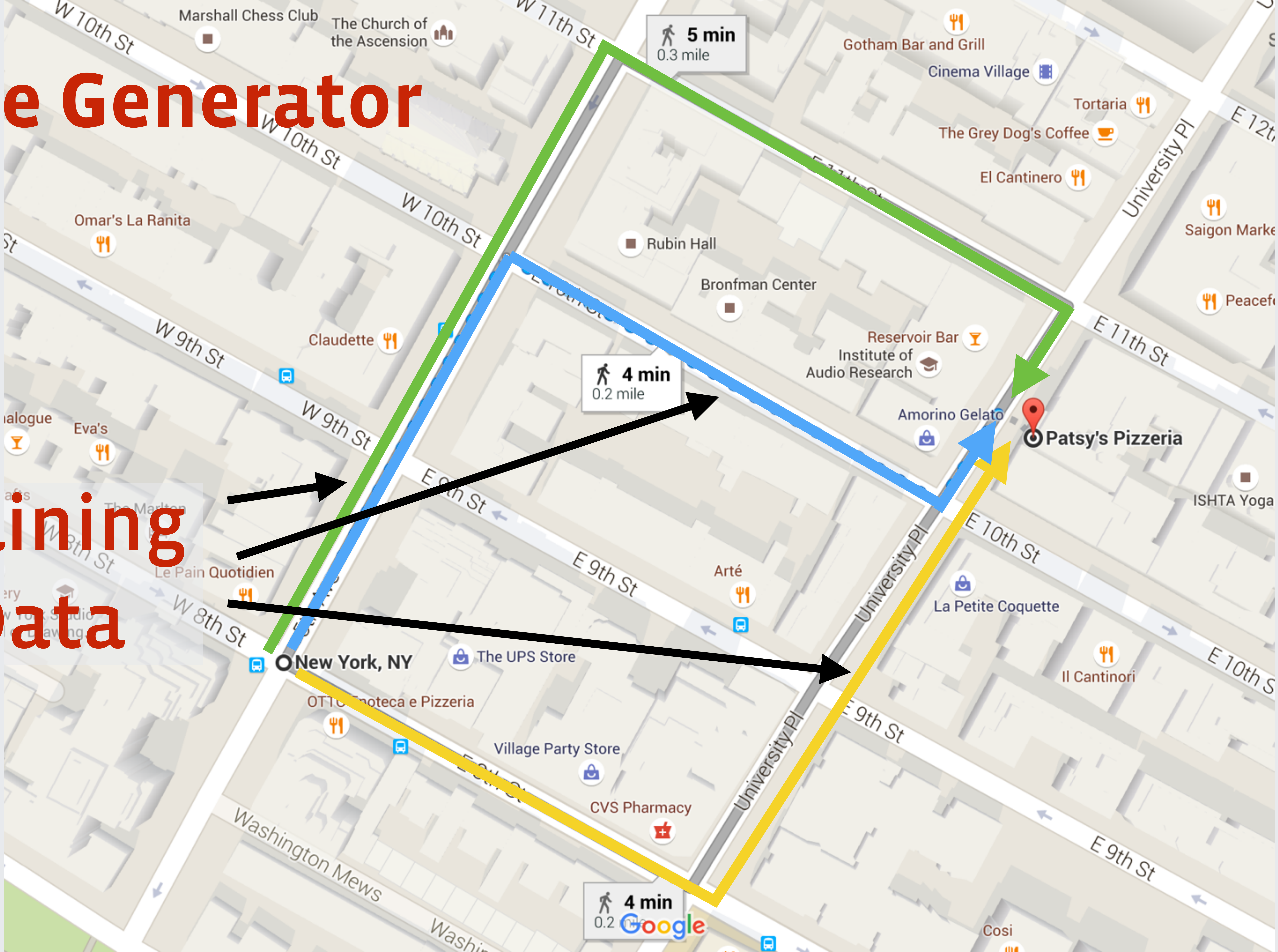
Route Generator

Linear Regressor with Mean-square Error



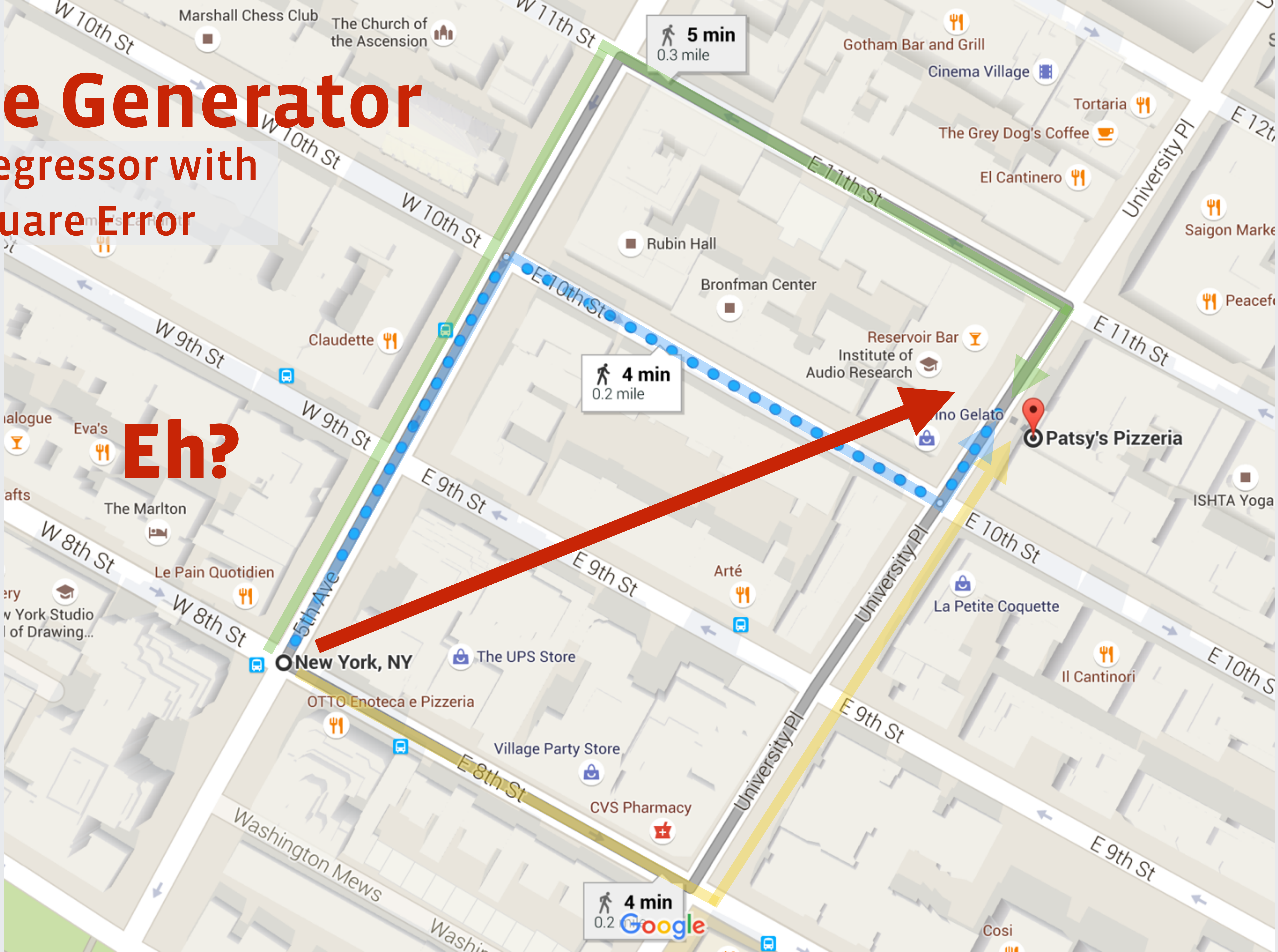
Route Generator

Training Data



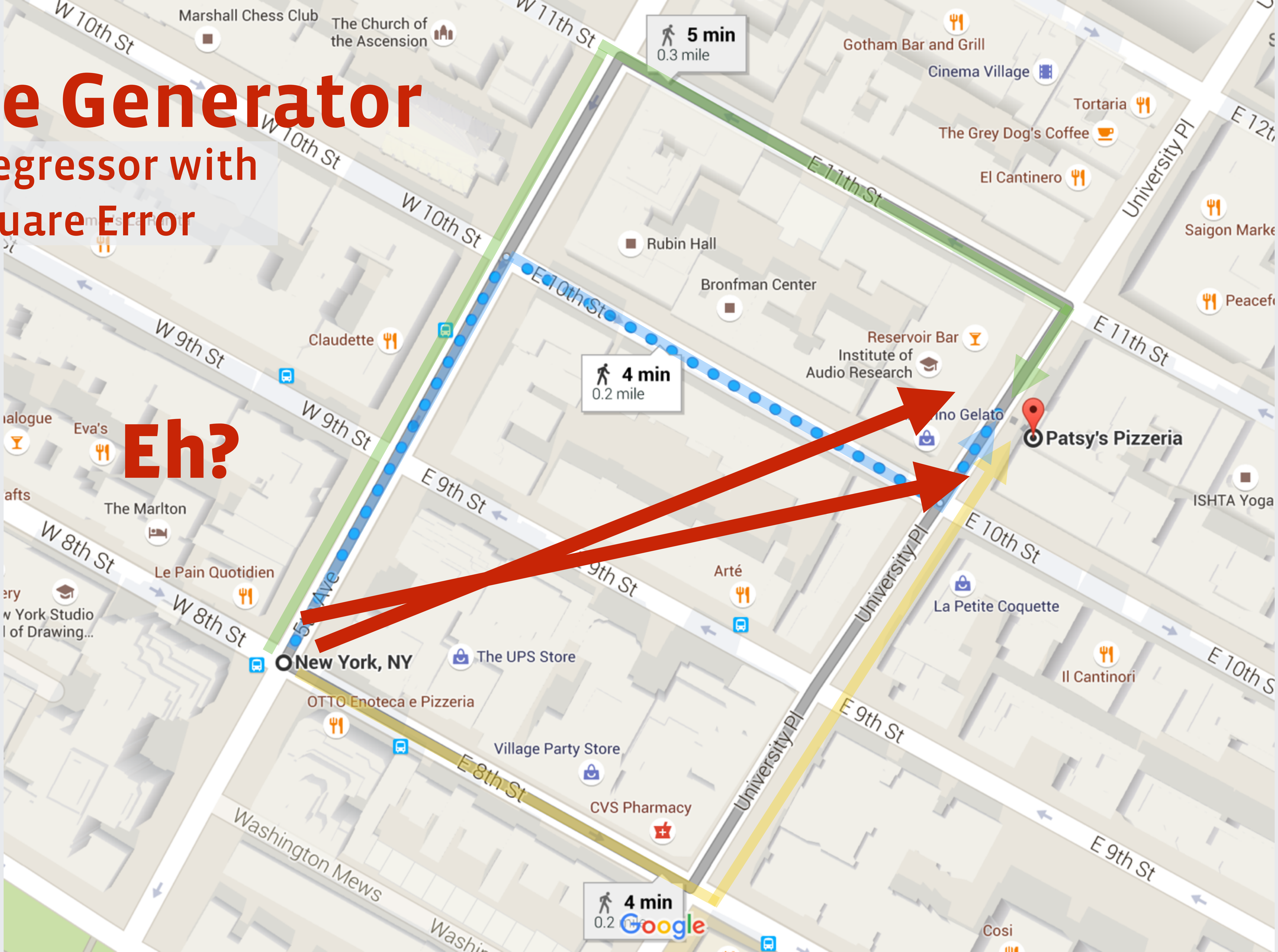
Route Generator

Linear Regressor with Mean-square Error



Route Generator

Linear Regressor with Mean-square Error



Route Generator

Linear Regressor with Mean-square Error

Route Generator

Linear Regressor with Mean-square Error

Route Generator

Linear Regressor with Mean-square Error

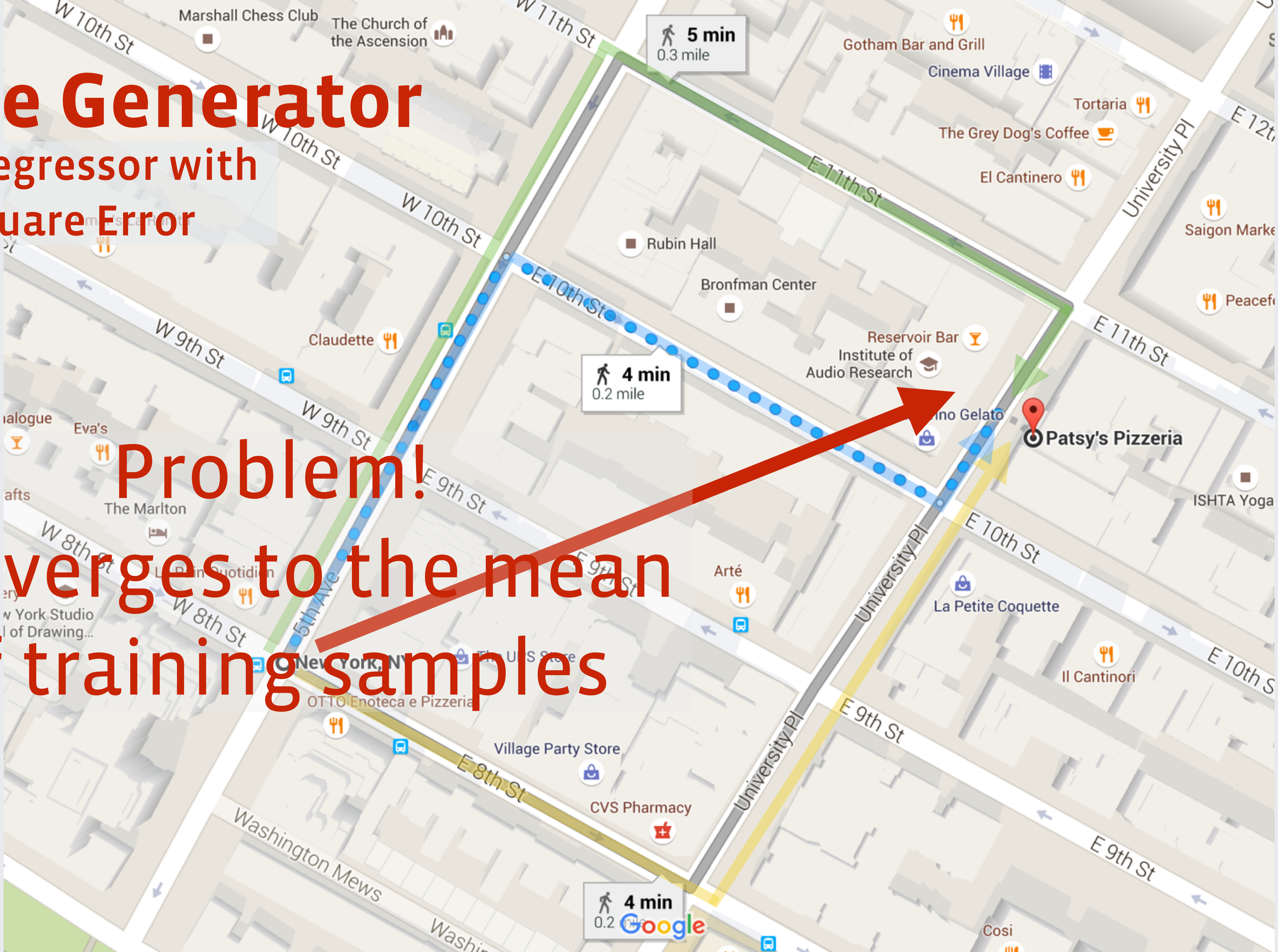
Route Generator

Linear Regressor with Mean-square Error

Eh? Eh? Eh? Eh? Eh?

Route Generator

Linear Regressor with Mean-square Error

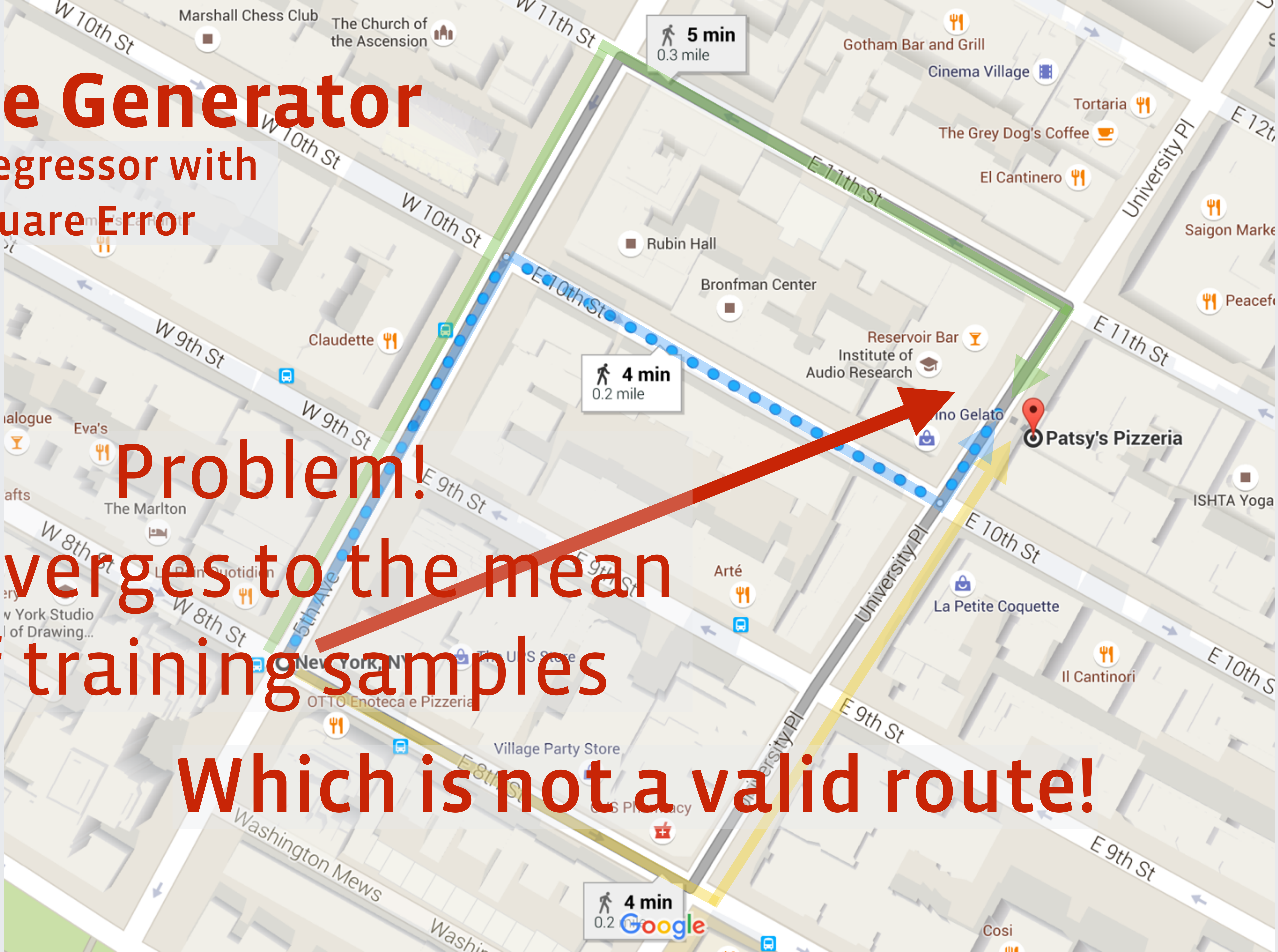


Problem!

Converges to the mean of training samples

Route Generator

Linear Regressor with Mean-square Error



Problem!

Converges to the mean of training samples

Which is not a valid route!

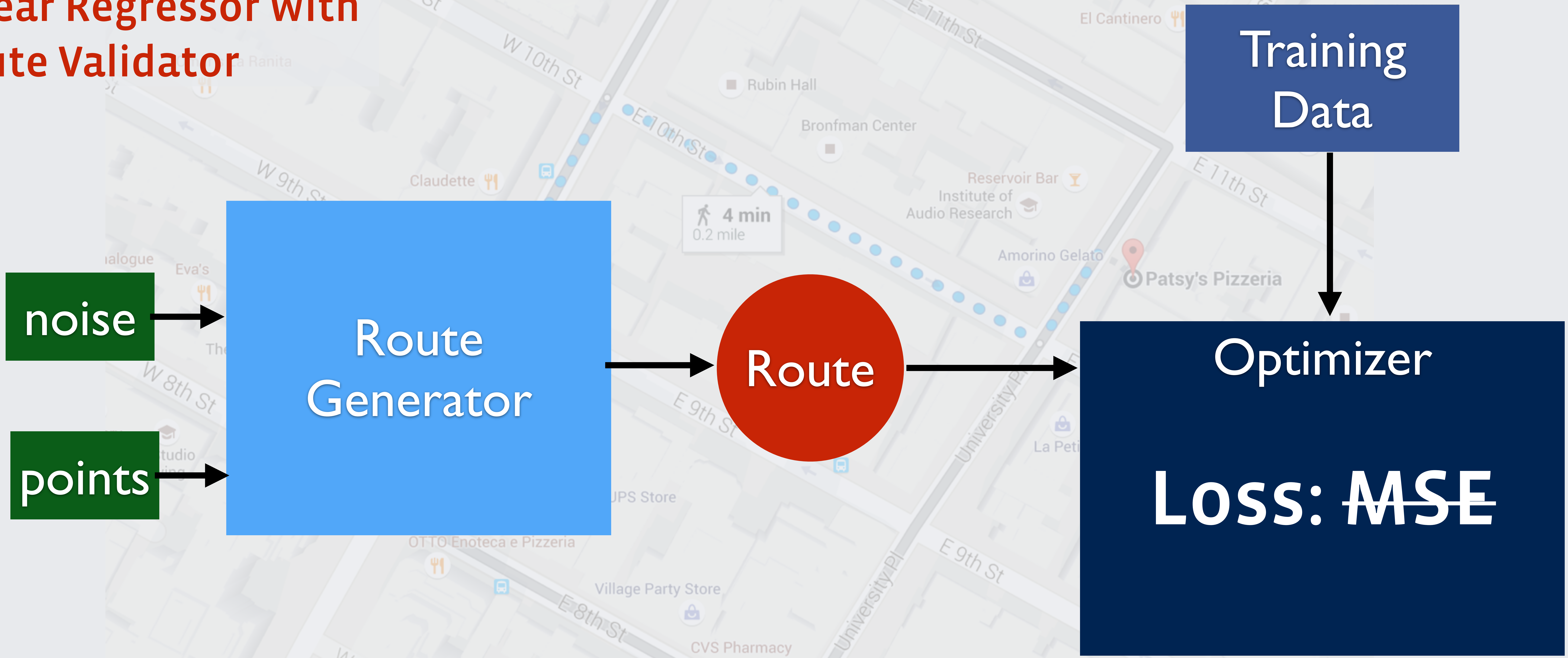
Route Generator

~~Linear Regressor with Mean-square Error~~

Let’s try again!

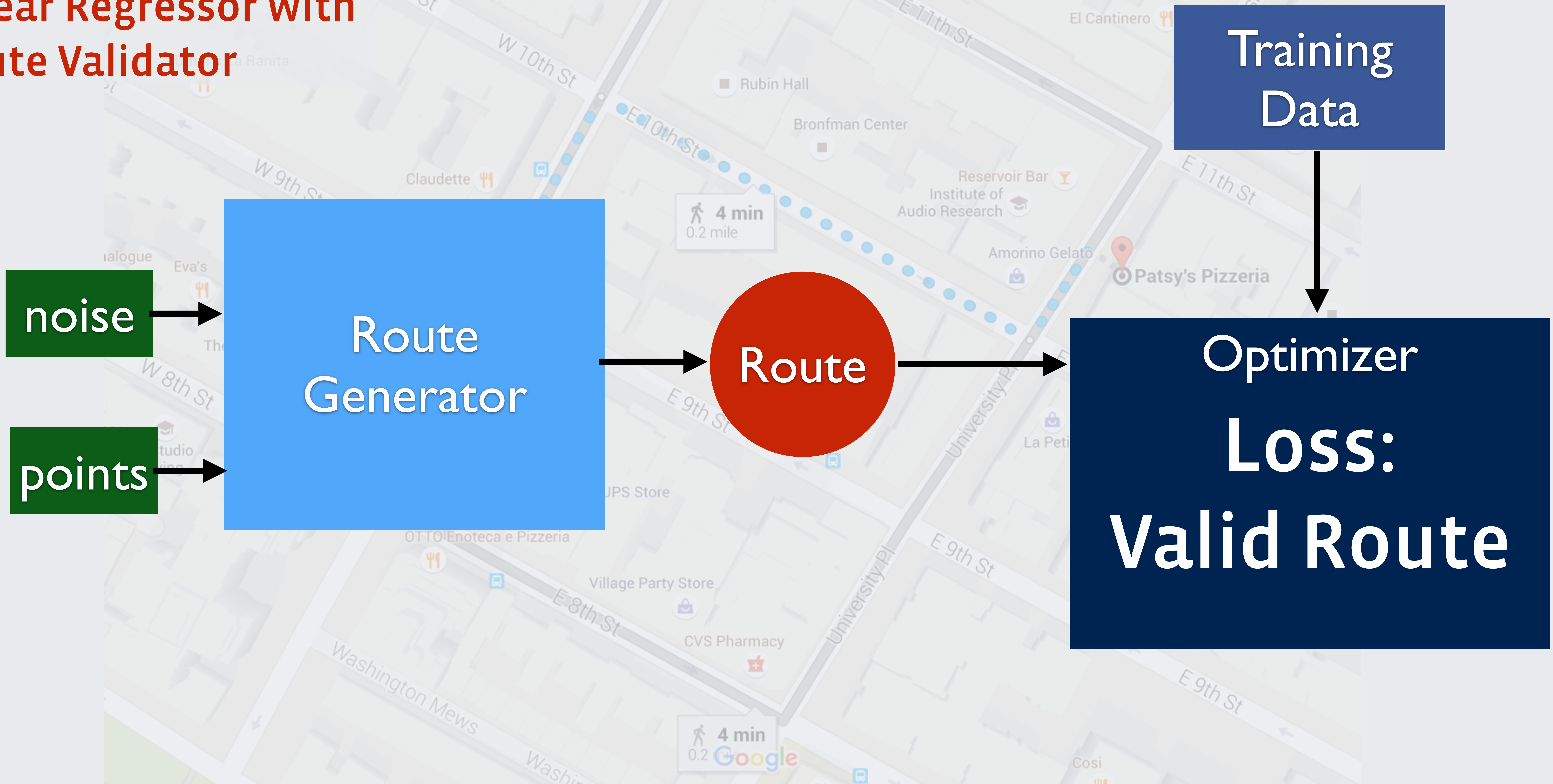
Route Generator

Linear Regressor with Route Validator



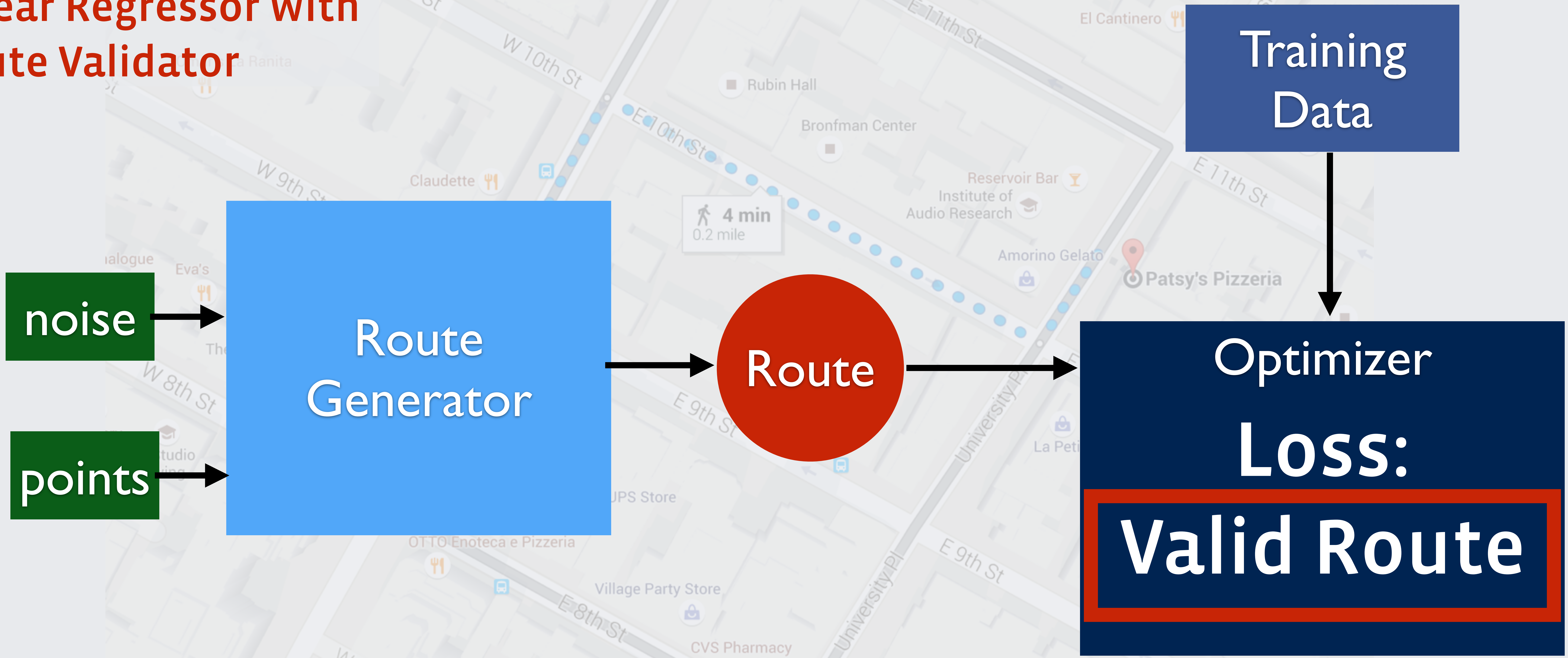
Route Generator

Linear Regressor with Route Validator



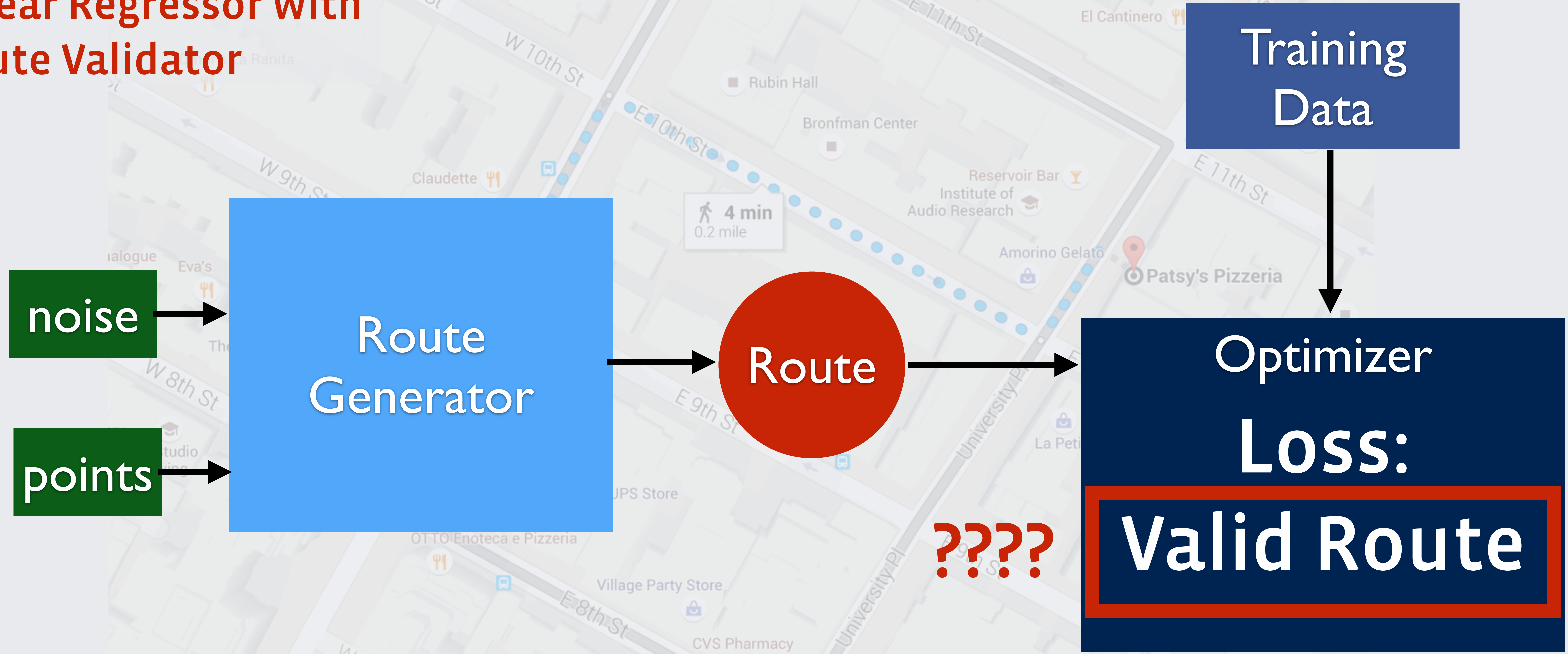
Route Generator

Linear Regressor with Route Validator

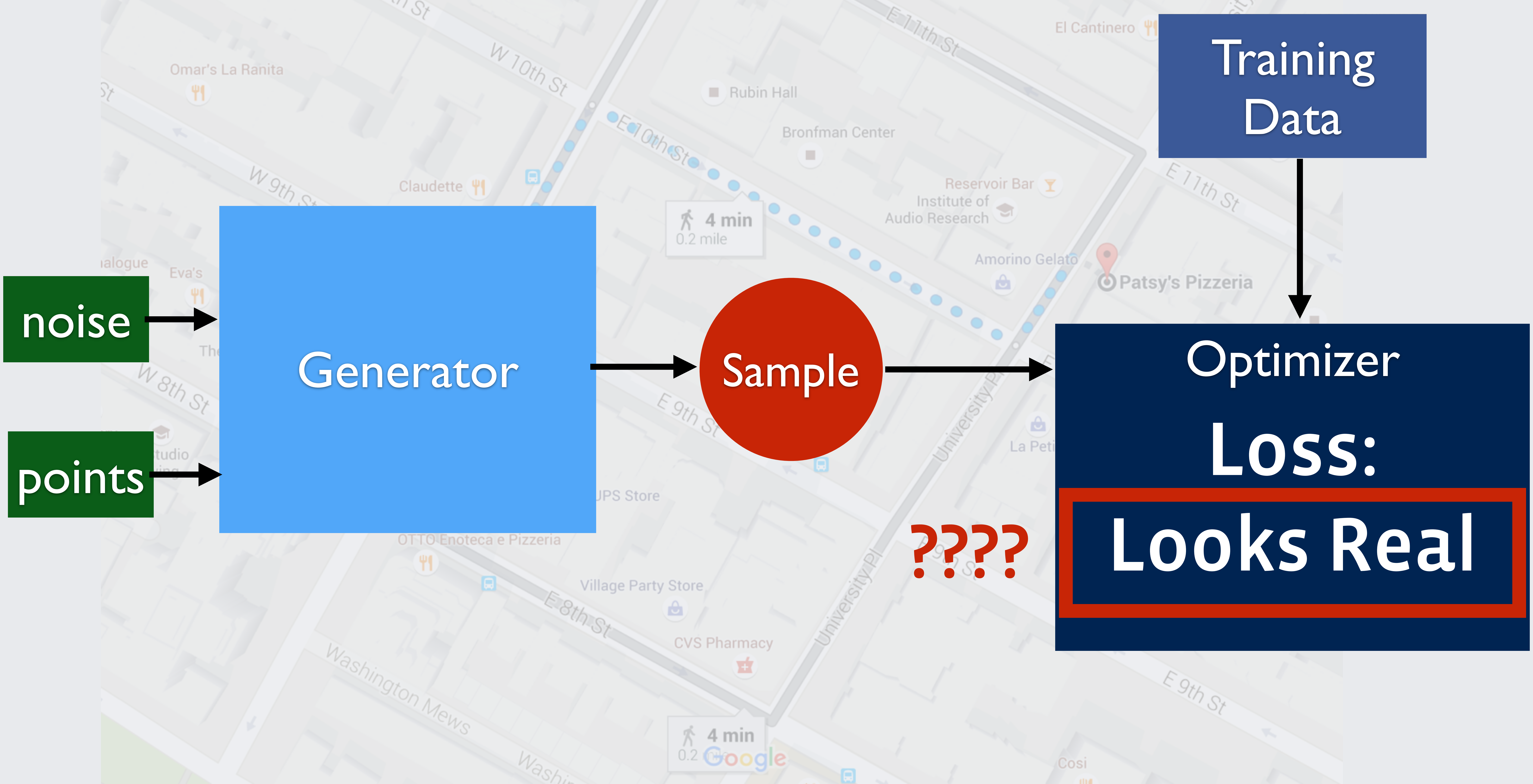


Route Generator

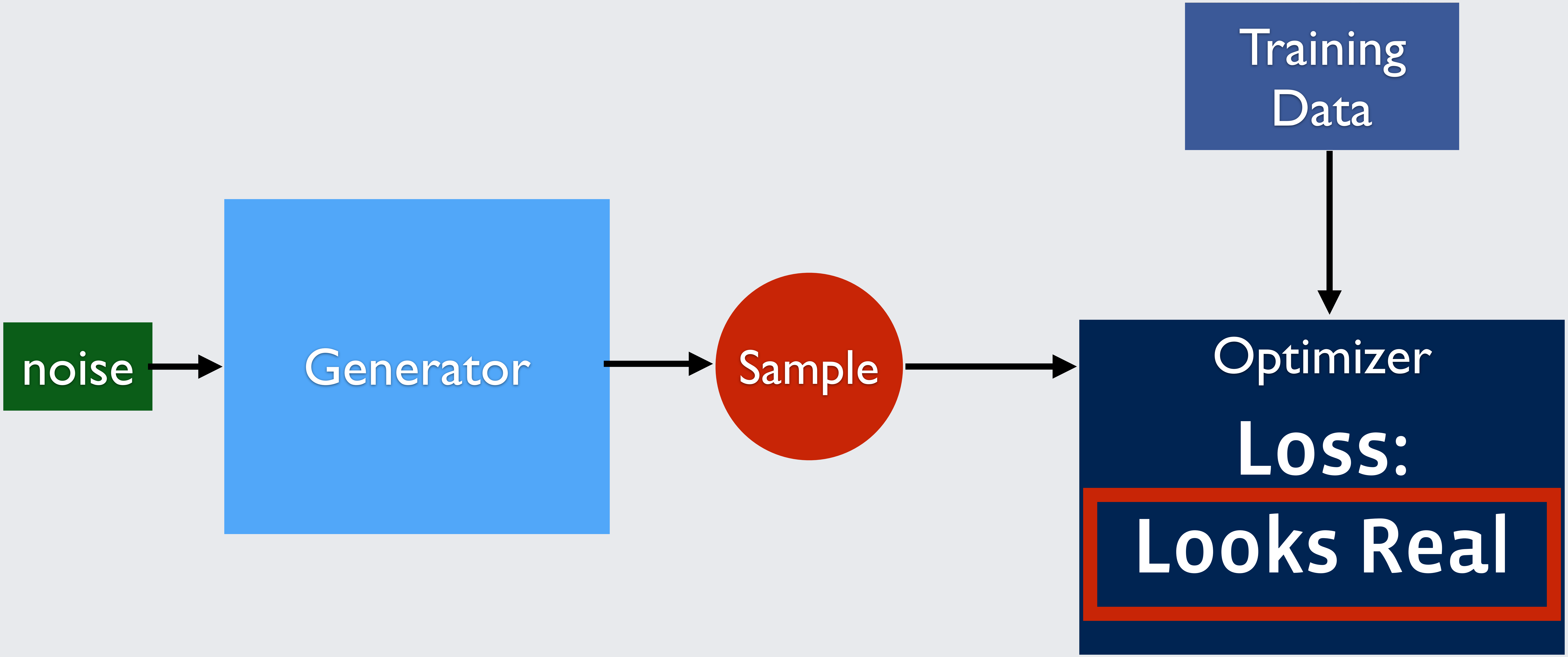
Linear Regressor with Route Validator



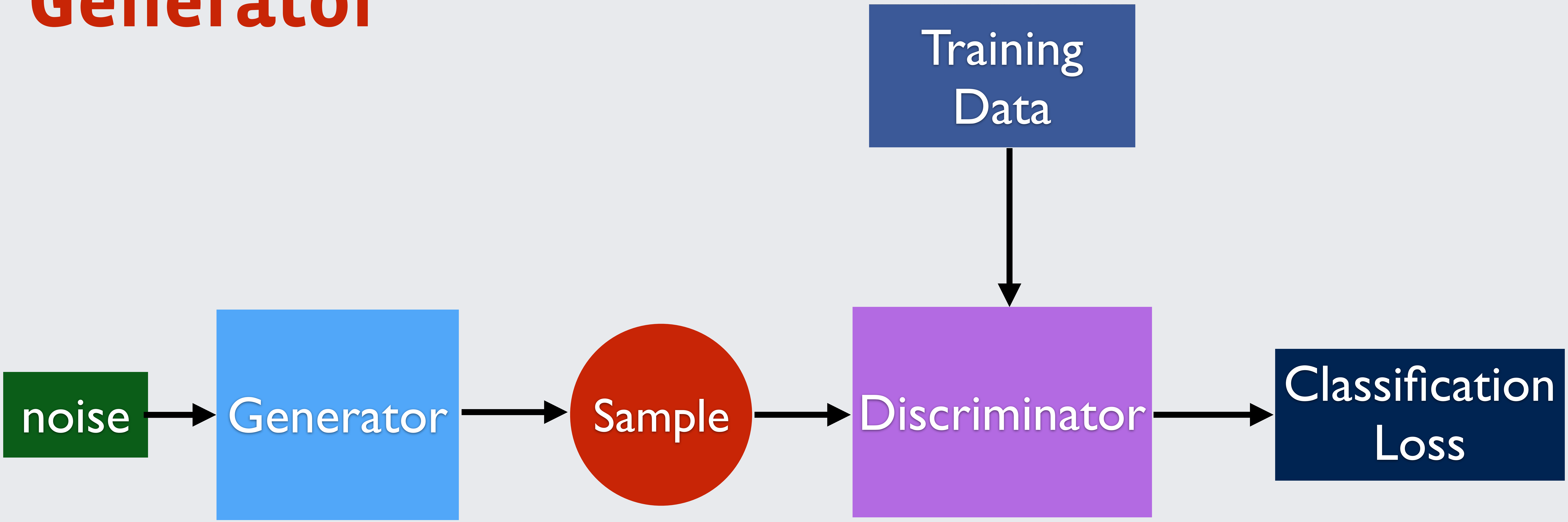
Generator


noise
points
Generator
Sample
Training Data
Optimizer
Loss:
Looks Real
????

Generator

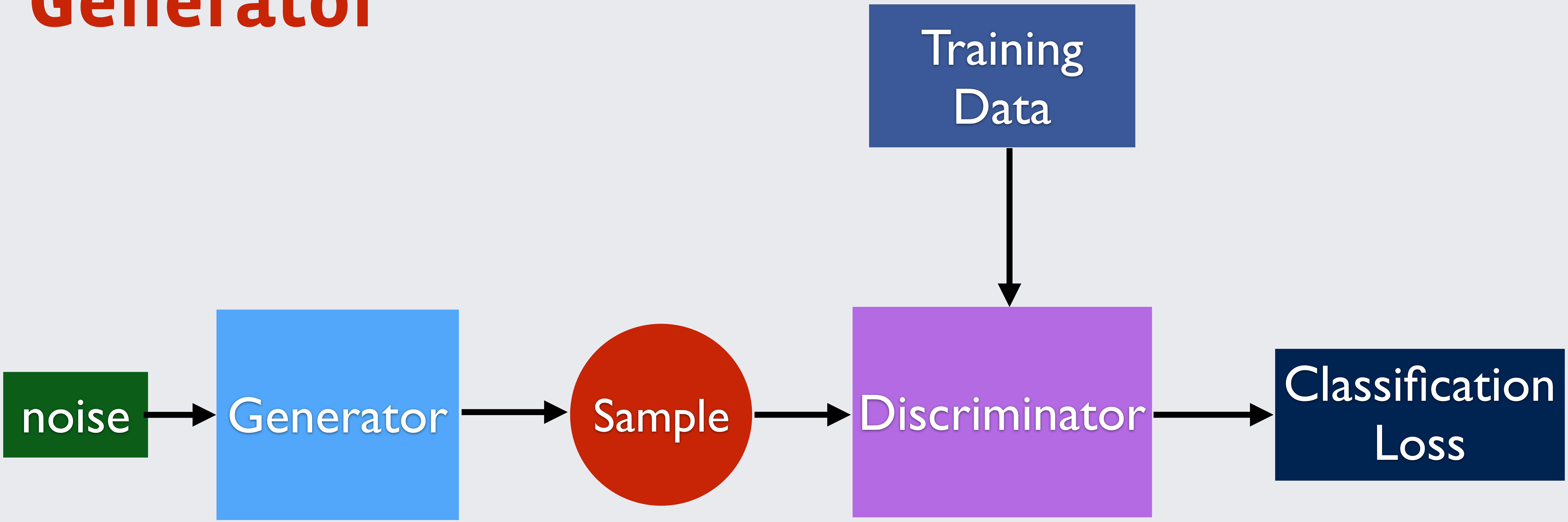

Training Data
noise
Generator
Sample
Optimizer
Loss:
Looks Real

Generator



Learnt Real/Fake Cost function

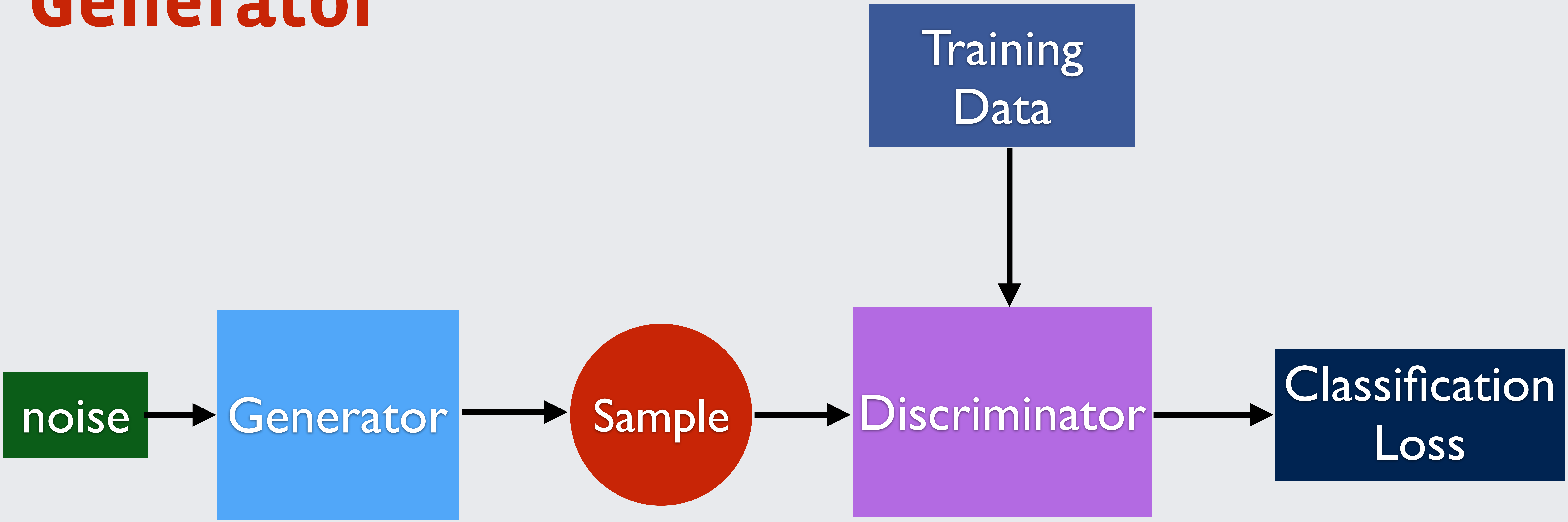
Generator



Neural Net

Neural Net

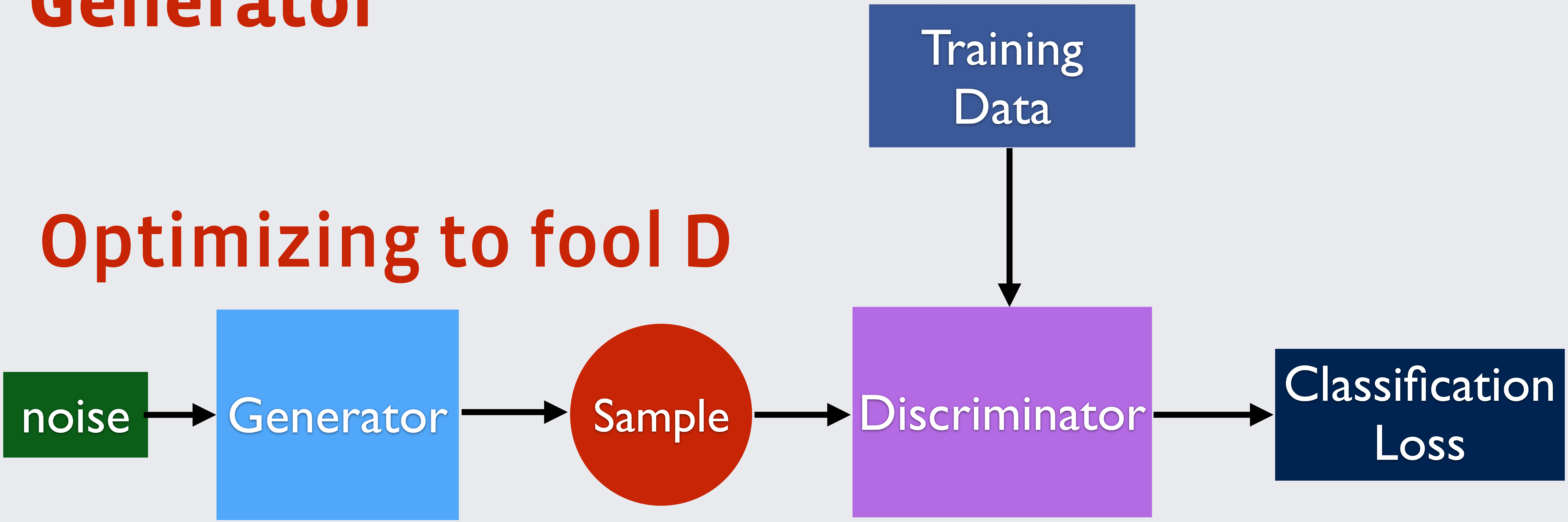
Generator



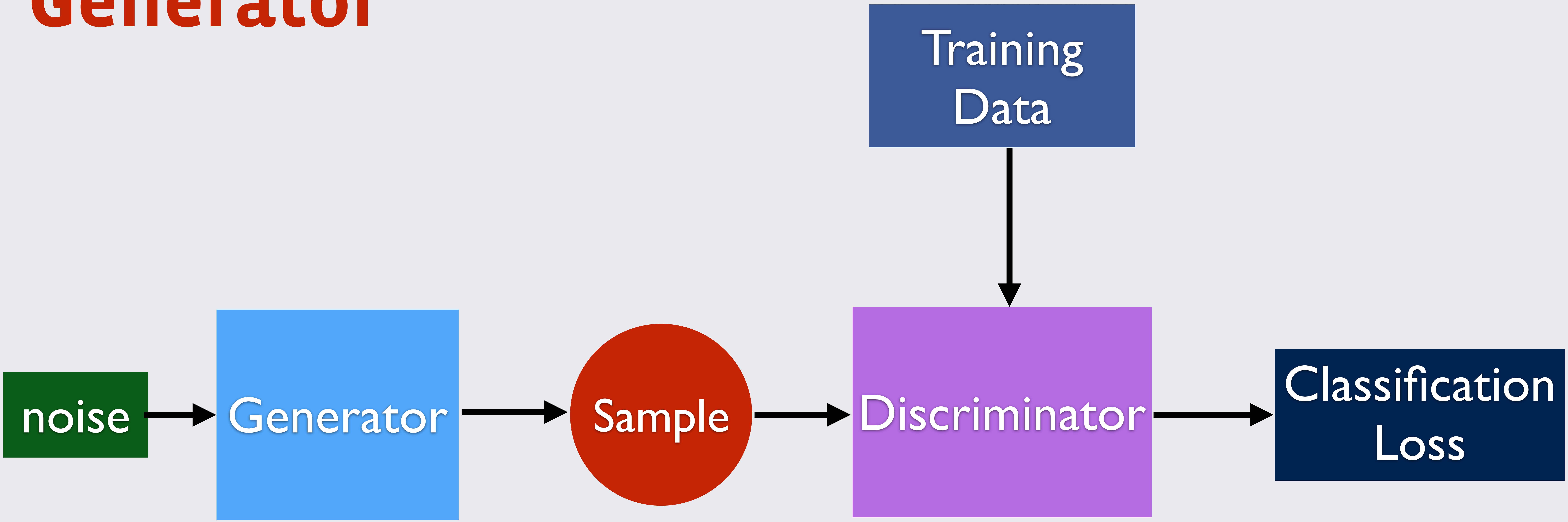
Trained via Gradient Descent

Generator

Optimizing to fool D



Generator



Optimizing to not get fooled by G

arxiv.org/abs/1406.2661

Cornell University Library

We

arXiv.org > stat > arXiv:1406.2661

Search or

Statistics > Machine Learning

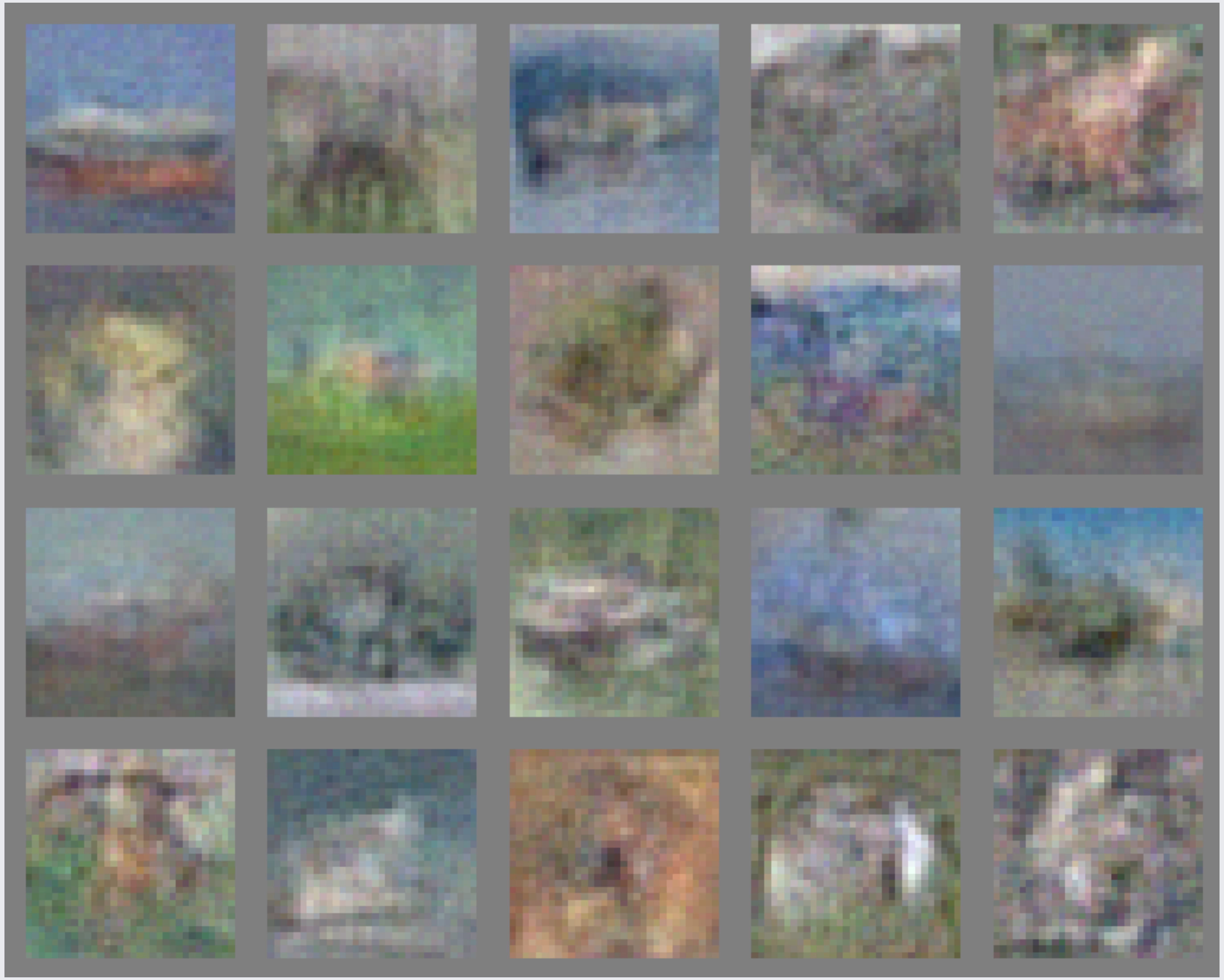
Generative Adversarial Networks

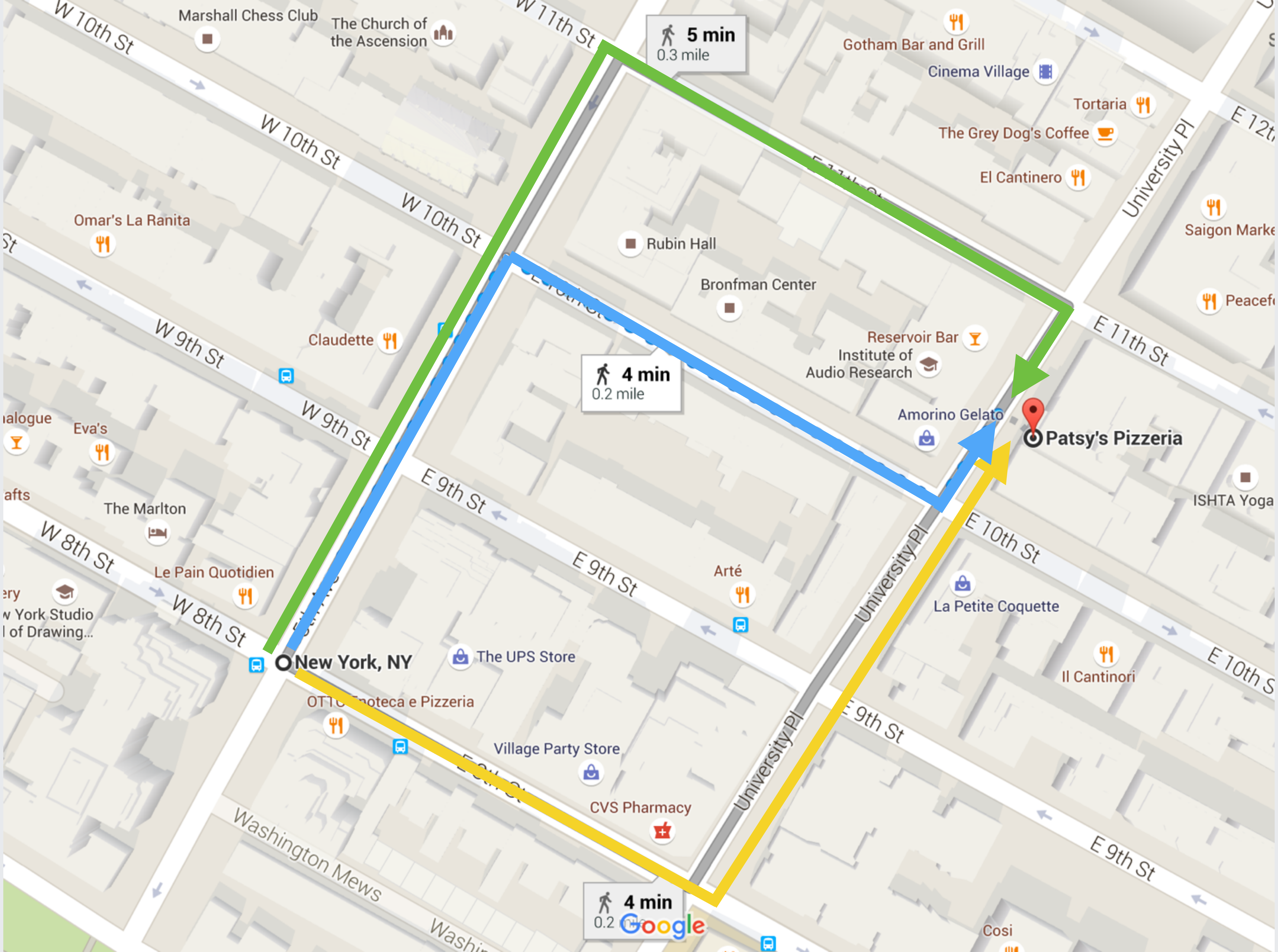
Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio

(Submitted on 10 Jun 2014)

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to 1/2 everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples. Experiments demonstrate the potential of the framework through qualitative and quantitative evaluation of the generated samples.

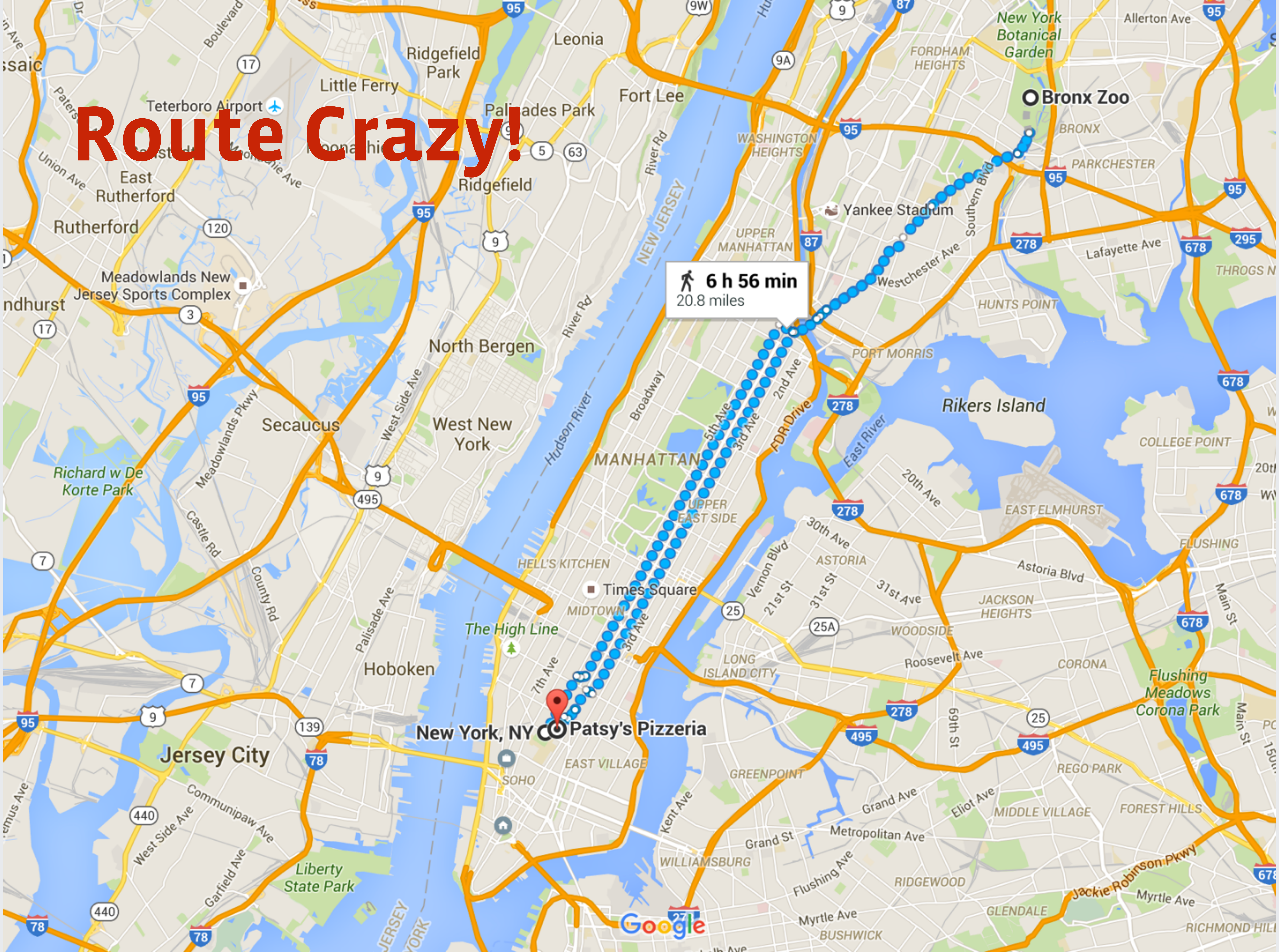
Subjects: **Machine Learning (stat.ML)**; Learning (cs.LG)
Cite as: arXiv:1406.2661 **[stat.ML]**
(or arXiv:1406.2661v1 **[stat.ML]** for this version)



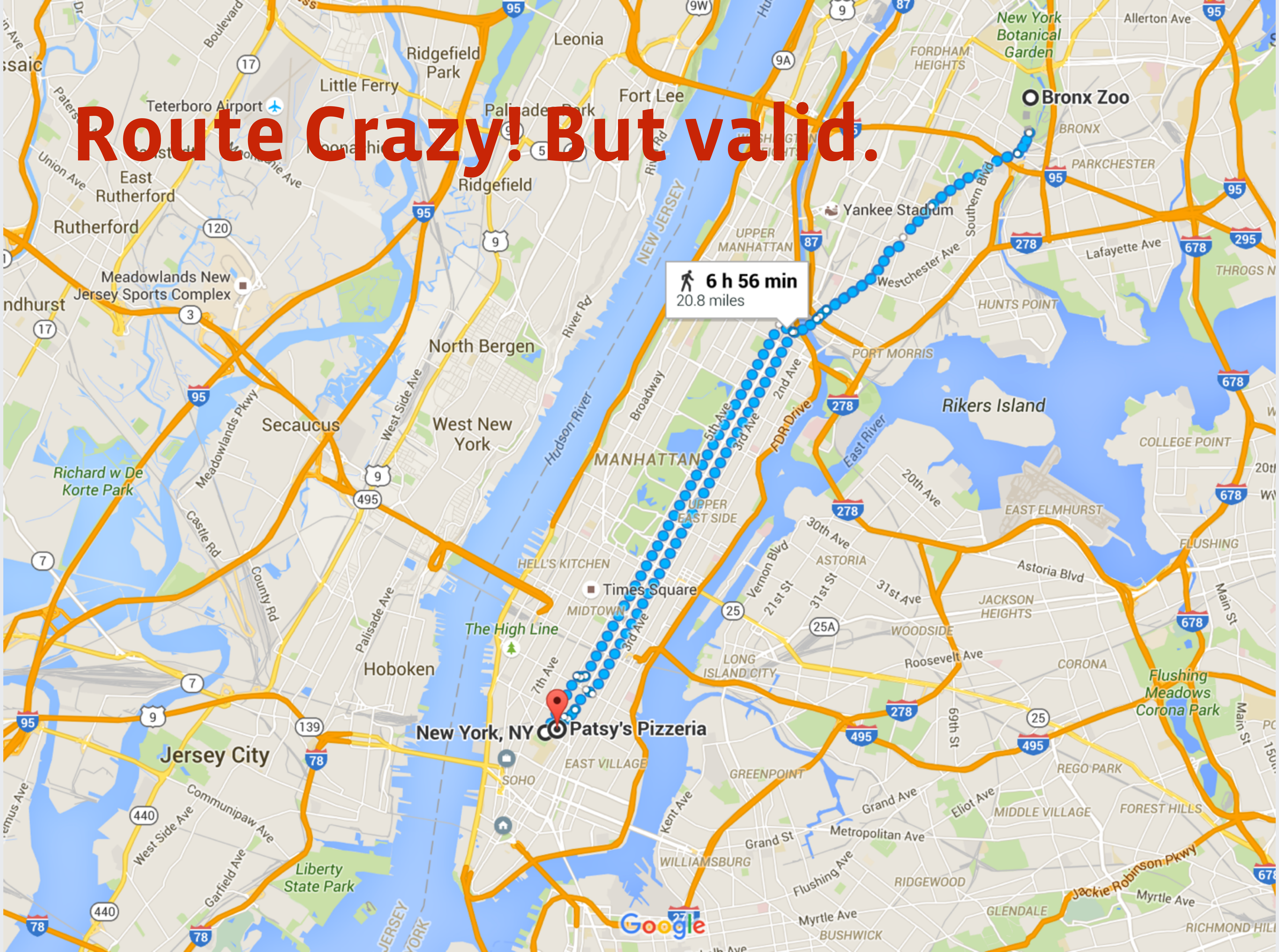

5 min
0.3 mile
4 min
0.2 mile
4 min
0.2
Patsy's Pizzeria
New York, NY
Marshall Chess Club
The Church of the Ascension
Gotham Bar and Grill
Cinema Village
Tortaria
The Grey Dog's Coffee
El Cantinero
Saigon Marke
Peacef
Omar's La Ranita
Rubin Hall
Bronfman Center
Claudette
Reservoir Bar
Institute of Audio Research
Amorino Gelato
ISHTA Yoga
Eva's
The Marlton
Le Pain Quotidien
Arté
La Petite Coquette
Il Cantinori
The UPS Store
OTTO Enoteca e Pizzeria
Village Party Store
CVS Pharmacy
Cosi
W 10th St
W 11th St
W 9th St
W 8th St
E 11th St
E 10th St
E 9th St
E 12th
University Pl
Washington Mews
Google

Route 4
Marshall Chess Club
The Church of the Ascension
Gotham Bar and Grill
Cinema Village
Tortaria
The Grey Dog's Coffee
El Cantinero
Omar's La Ranita
Rubin Hall
Bronfman Center
Claudette
Reservoir Bar
Institute of Audio Research
Patsy's Pizzeria
Eva's
The Marlton
Le Pain Quotidien
Arté
La Petite Coquette
ISHTA Yoga
New York, NY
The UPS Store
OTTO Enoteca e Pizzeria
Il Cantinori
Village Party Store
CVS Pharmacy
Washington Mews
Cosi
W 10th St
W 11th St
5th Ave
E 11th St
E 10th St
W 9th St
E 9th St
W 8th St
E 8th St
University Pl
Google

Route Crazy!


6 h 56 min
20.8 miles
Bronx Zoo
Patsy's Pizzeria
New York, NY
Google

Route Crazy! But valid.



arxiv.org/abs/1506.05751

Cornell University Library

arXiv.org > cs > arXiv:1506.05751

Search

Computer Science > Computer Vision and Pattern Recognition

Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks

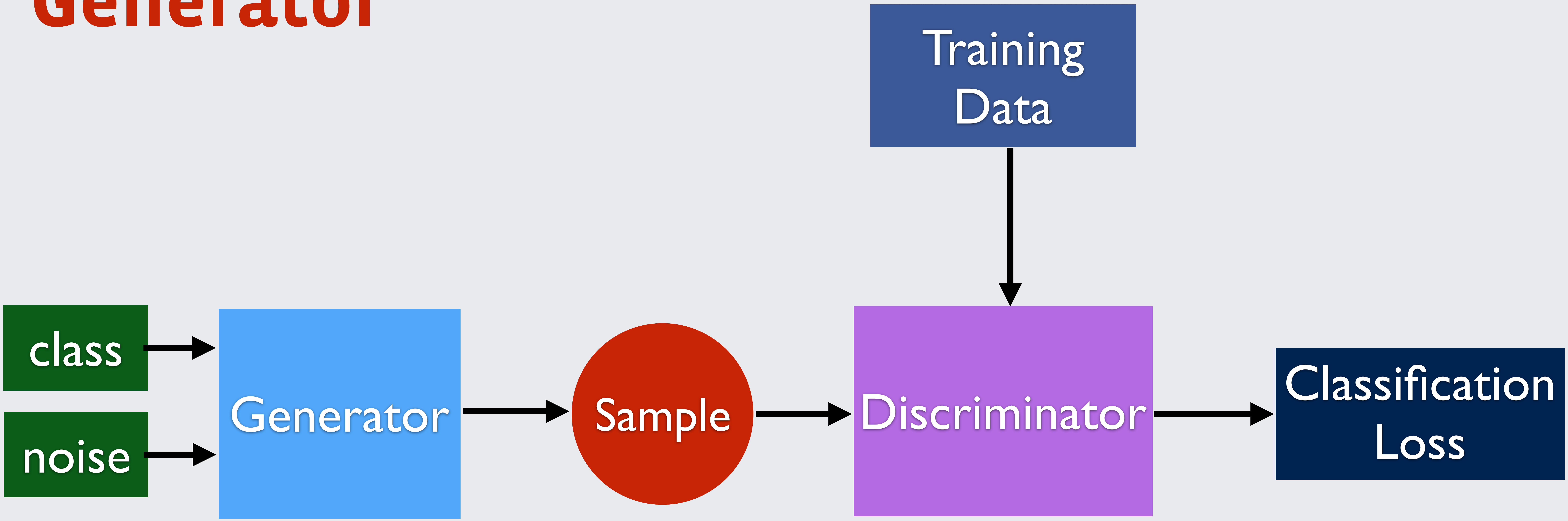
Emily Denton, Soumith Chintala, Arthur Szlam, Rob Fergus

(Submitted on 18 Jun 2015)

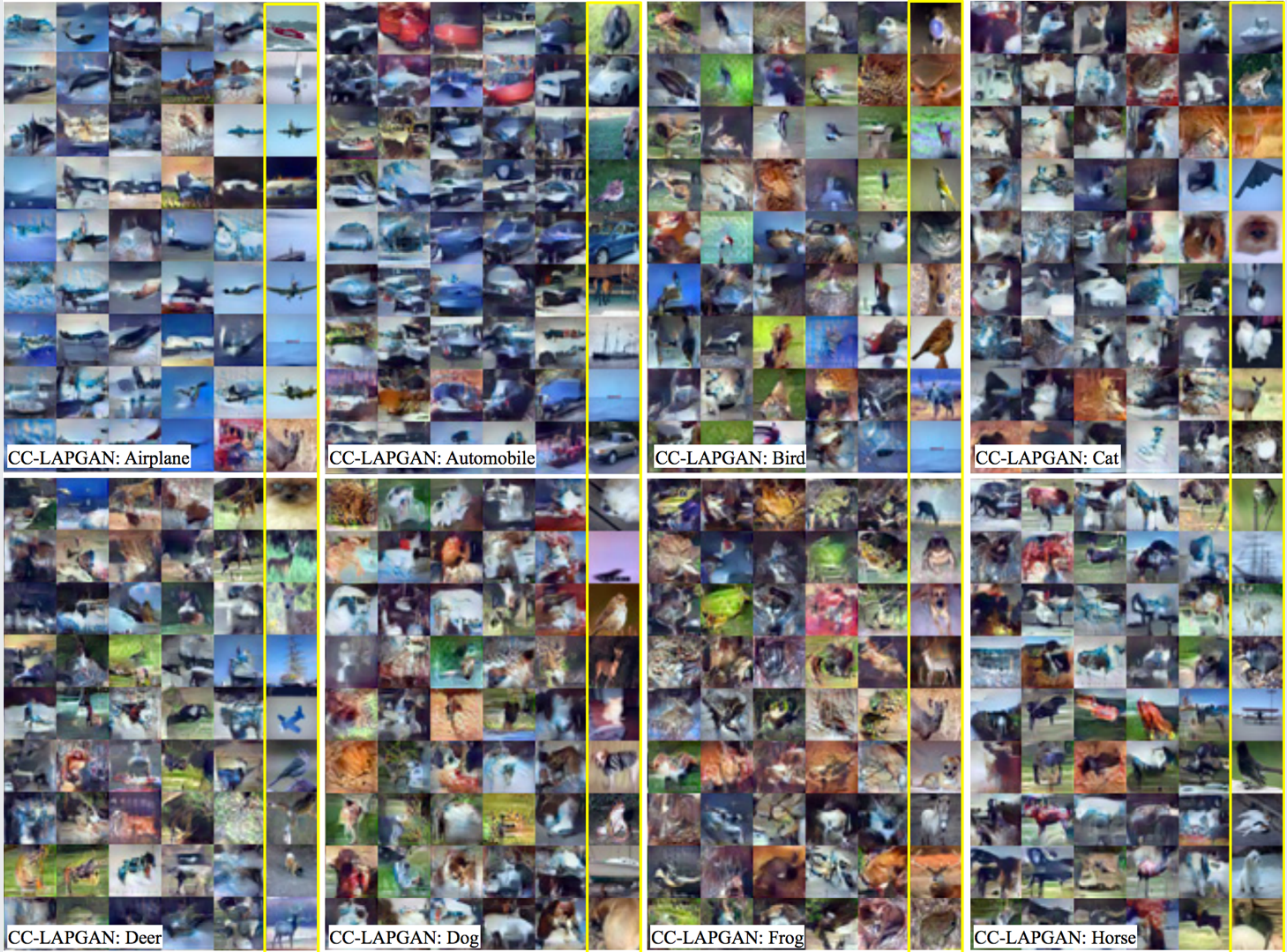
In this paper we introduce a generative parametric model capable of producing high quality samples of natural images. Our approach uses a cascade of convolutional networks within a Laplacian pyramid framework to generate images in a coarse-to-fine fashion. At each level of the pyramid, a separate generative convnet model is trained using the Generative Adversarial Nets (GAN) approach (Goodfellow et al.). Samples drawn from our model are of significantly higher quality than alternate approaches. In a quantitative assessment by human evaluators, our CIFAR10 samples were mistaken for real images around 40% of the time, compared to 10% for samples drawn from a GAN baseline model. We also show samples from models trained on the higher resolution images of the LSUN scene dataset.

Subjects: **Computer Vision and Pattern Recognition (cs.CV)**
Cite as: arXiv:1506.05751 [cs.CV]
(or arXiv:1506.05751v1 [cs.CV] for this version)

Generator



Optimizing to not get fooled by G


CC-LAPGAN: Airplane
CC-LAPGAN: Automobile
CC-LAPGAN: Bird
CC-LAPGAN: Cat
CC-LAPGAN: Deer
CC-LAPGAN: Dog
CC-LAPGAN: Frog
CC-LAPGAN: Horse

arxiv.org/abs/1511.05440

Cornell University Library

We

arXiv.org > cs > arXiv:1511.05440

Search or

Computer Science > Learning

Deep multi-scale video prediction beyond mean square error

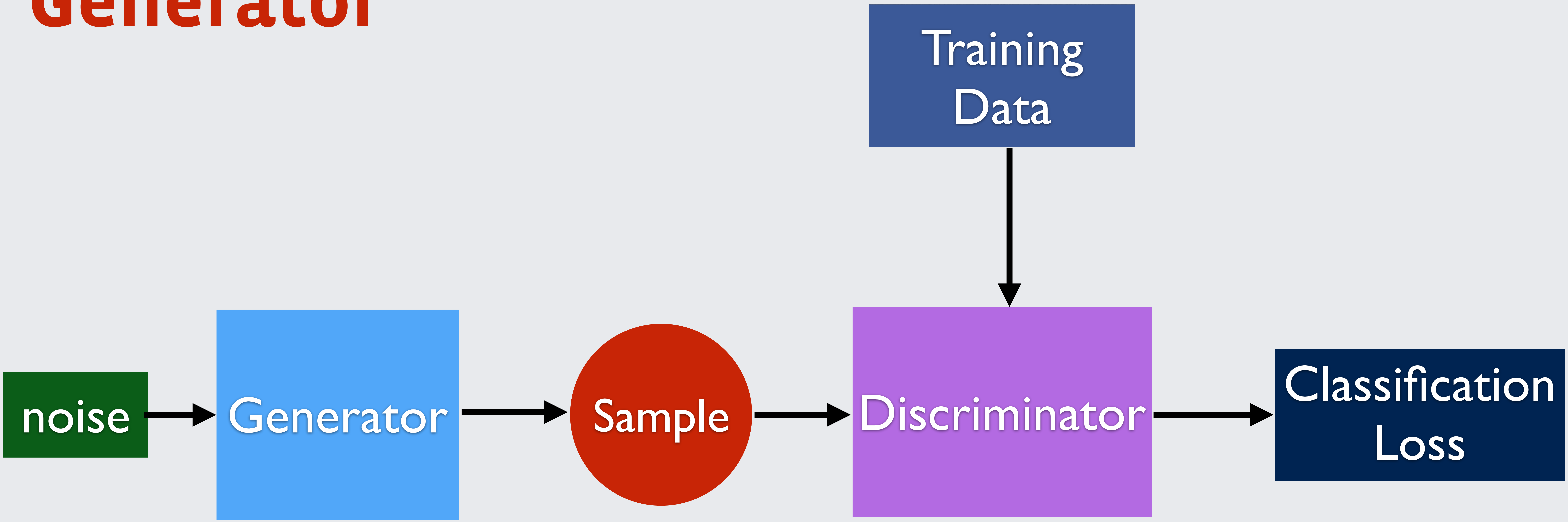
Michael Mathieu, Camille Couprie, Yann LeCun

(Submitted on 17 Nov 2015 (v1), last revised 26 Feb 2016 (this version, v6))

Learning to predict future images from a video sequence involves the construction of an internal representation that models the image evolution accurately, and therefore, to some degree, its content and dynamics. This is why pixel-space video prediction may be viewed as a promising avenue for unsupervised feature learning. In addition, while optical flow has been a very studied problem in computer vision for a long time, future frame prediction is rarely approached. Still, many vision applications could benefit from the knowledge of the next frames of videos, that does not require the complexity of tracking every pixel trajectories. In this work, we train a convolutional network to generate future frames given an input sequence. To deal with the inherently blurry predictions obtained from the standard Mean Squared Error (MSE) loss function, we propose three different and complementary feature learning strategies: a multi-scale architecture, an adversarial training method, and an image gradient difference loss function. We compare our predictions to different published results based on recurrent neural networks on the UCF101 dataset

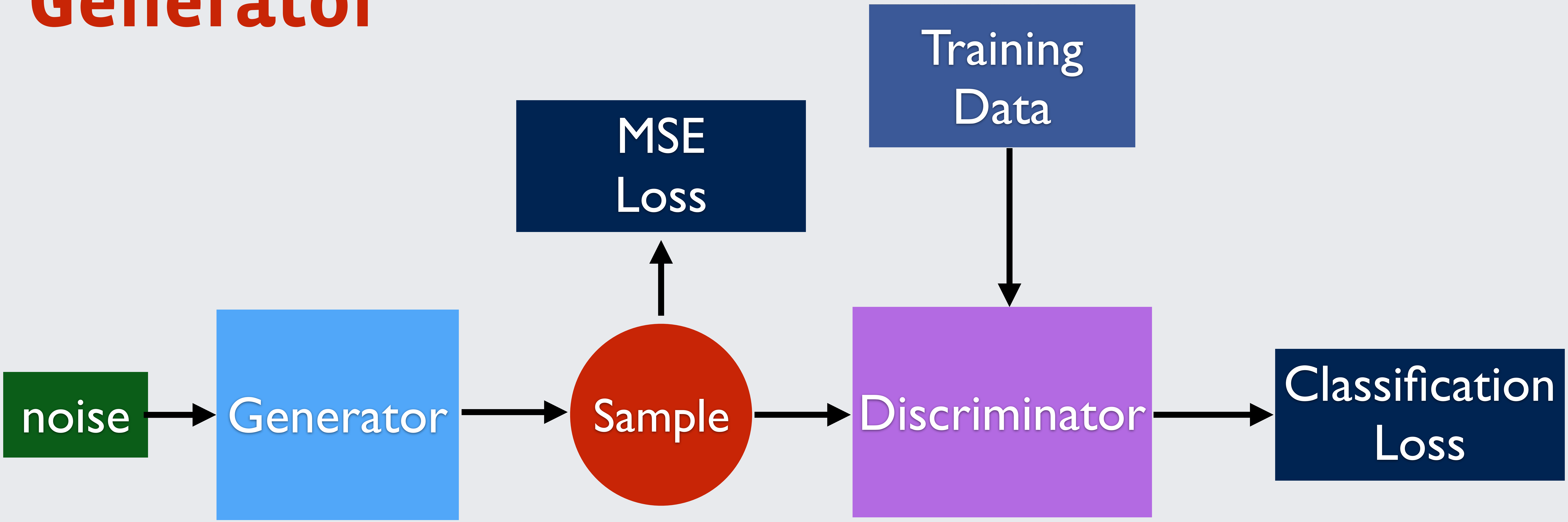
Subjects: **Learning (cs.LG)**; Computer Vision and Pattern Recognition (cs.CV); Machine Learning (stat.ML)
Cite as: arXiv:1511.05440 **[cs.LG]**
(or arXiv:1511.05440v6 **[cs.LG]** for this version)

Generator



Optimizing to not get fooled by G

Generator



Optimizing to not get fooled by G


Input frames
Ground truth
ℓ_2 result
ℓ_1 result
GDL ℓ_1 result
Adversarial result
Adversarial+GDL result

arxiv.org/abs/1511.06434

Cornell University Library

arXiv.org > cs > arXiv:1511.06434

Search

Computer Science > Learning

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks

Alec Radford, Luke Metz, Soumith Chintala

(Submitted on 19 Nov 2015 (v1), last revised 7 Jan 2016 (this version, v2))

In recent years, supervised learning with convolutional networks (CNNs) has seen huge adoption in computer vision applications. Comparatively, unsupervised learning with CNNs has received less attention. In this work we hope to help bridge the gap between the success of CNNs for supervised learning and unsupervised learning. We introduce a class of CNNs called deep convolutional generative adversarial networks (DCGANs), that have certain architectural constraints, and demonstrate that they are a strong candidate for unsupervised learning. Training on various image datasets, we show convincing evidence that our deep convolutional adversarial pair learns a hierarchy of representations from object parts to scenes in both the generator and discriminator. Additionally, we use the learned features for novel tasks - demonstrating their applicability as general image representations.

Comments: Under review as a conference paper at ICLR 2016
Subjects: **Learning (cs.LG)**; Computer Vision and Pattern Recognition (cs.CV)
Cite as: arXiv:1511.06434 **[cs.LG]**
(or arXiv:1511.06434v2 **[cs.LG]** for this version)

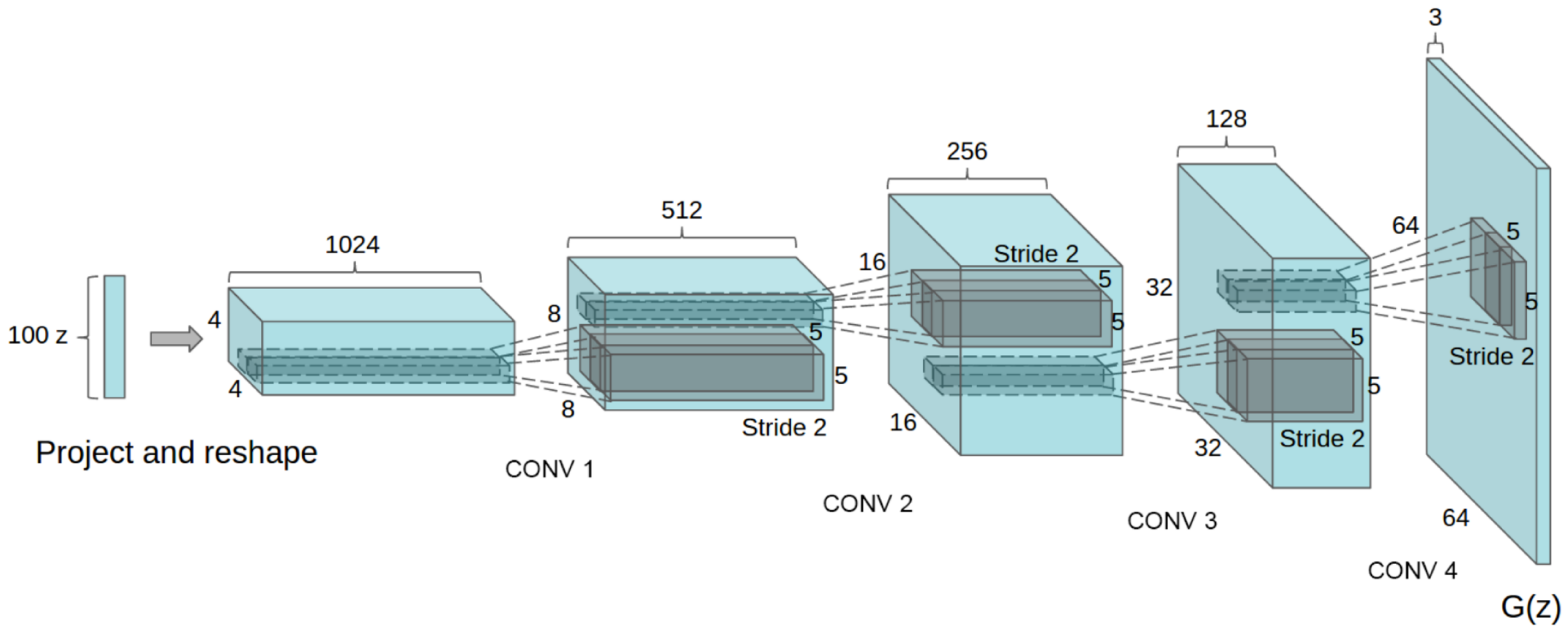
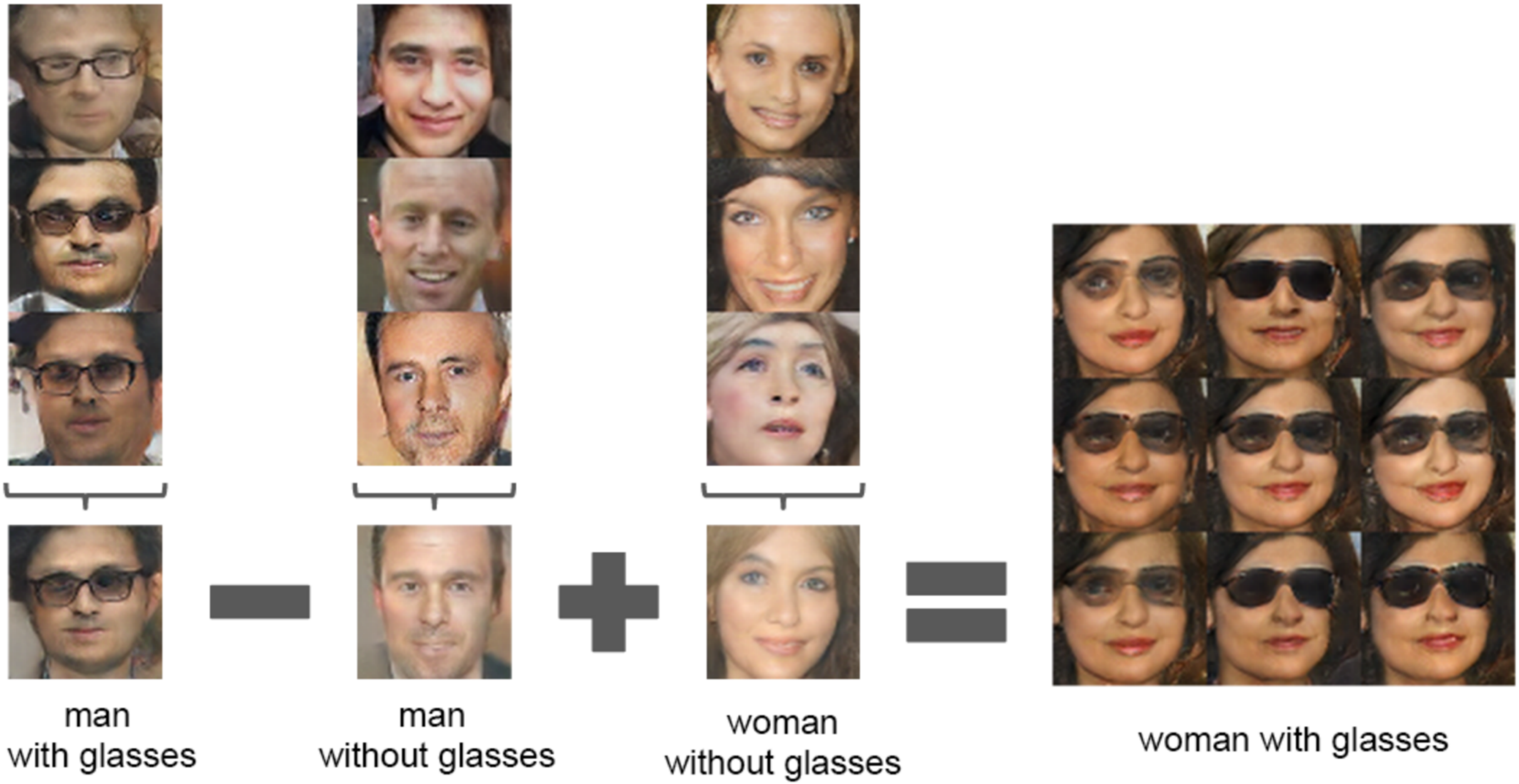

100 z
Project and reshape
1024
4
4
512
8
8
Stride 2
5
5
CONV 1
256
16
16
Stride 2
5
5
CONV 2
128
32
32
Stride 2
5
5
CONV 3
3
64
64
5
5
Stride 2
CONV 4
G(z)

Table 2: SVHN classification with 1000 labels

Model	error rate
KNN	77.93%
TSVM	66.55%
M1+KNN	65.63%
M1+TSVM	54.33%
M1+M2	36.02%
SWWAE without dropout	27.83%
SWWAE with dropout	23.56%
DCGAN (ours) + L2-SVM	22.48%
Supervised CNN with the same architecture	28.87% (validation)


man
with glasses
man
without glasses
woman
without glasses
woman with glasses

Uses

- Unsupervised Learning
  - Learn when there’s little labeled data
- Planning
  - Look-ahead to take better decisions