Mismatched training and dev/test data

Addressing data mismatch

deeplearning.ai

Addressing data mismatch

→ • Carry out manual error analysis to try to understand differe
between training and dev/test sets

E.g. noisy — car noise   street numbers

→ • Make training data more similar; or collect more data simila
dev/test sets

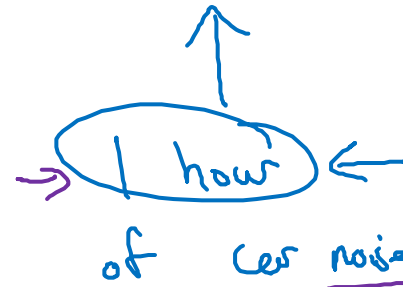E.g. Simulate noisy in-car data

Artificial data synthesis

"The quick brown
fox jumps
over the lazy dog."

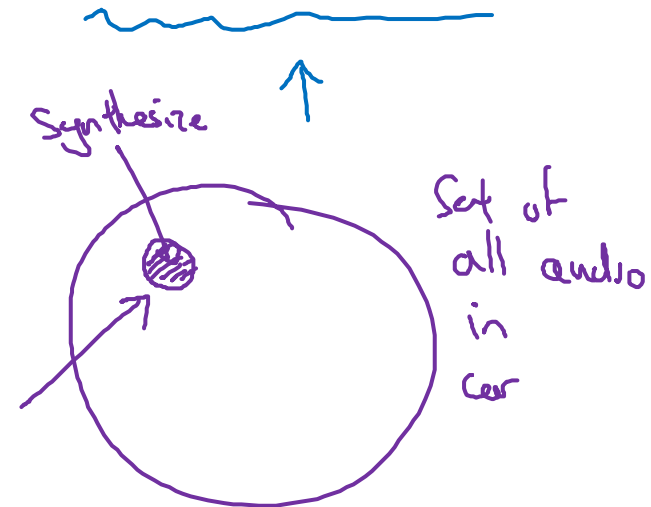+   Car noise   =   Synthesized
in-car audio

10,000 hours

1 hour
of car noise

Overfit to 1 hour of
car noise
10,000 hours

Synthesize

Set of
all audio
in
car

Andrew Ng

Artificial data synthesis

Car recognition:

≈ 20 cars

Synthesized → All cars

Andrew Ng