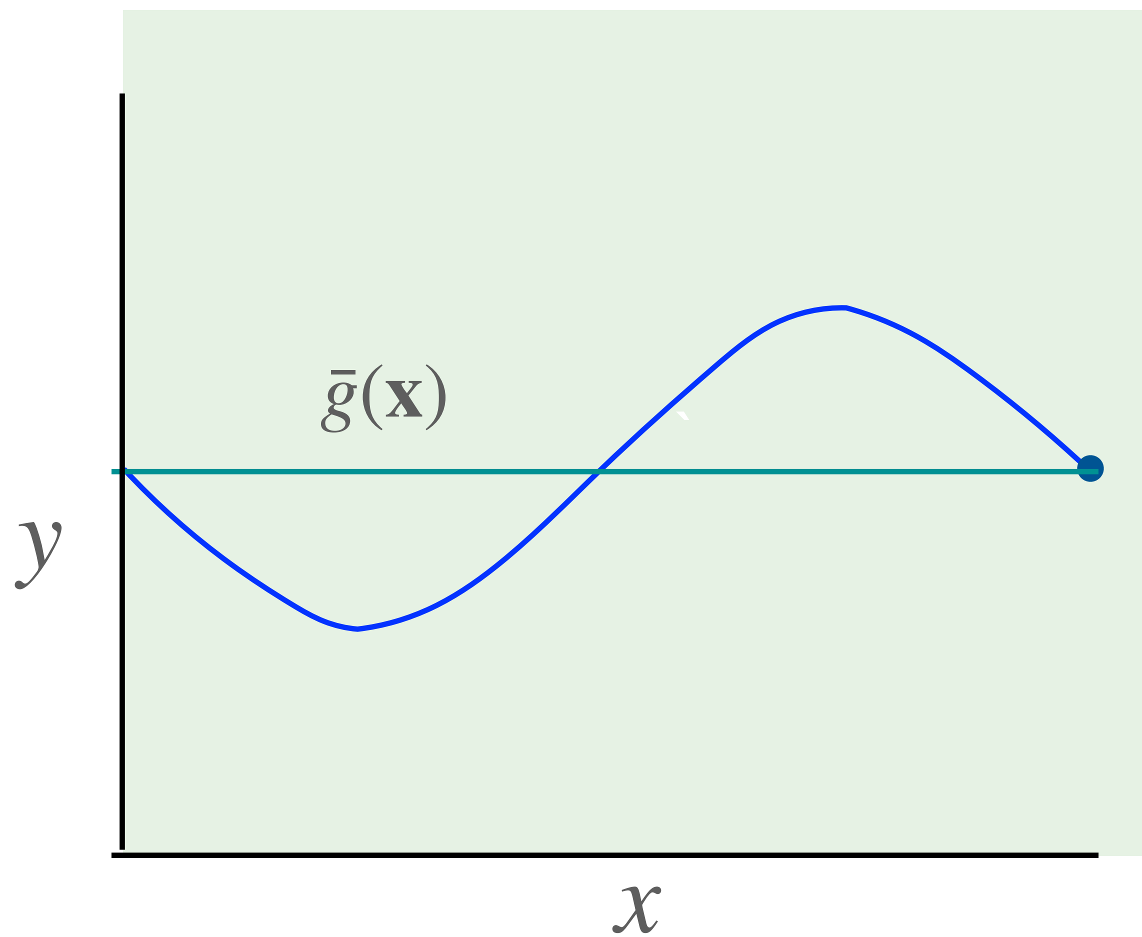
Notation and Math

Bias (also call "Bias squared") of a model/hypothesis class



For a fixed $\mathbf{x}$

$$\text{bias}(\mathbf{x}) = (f(\mathbf{x}) - \bar{g}(\mathbf{x}))^2$$

General case

$$\text{bias} = E_{\mathbf{x}}[(f(\mathbf{x}) - \bar{g}(\mathbf{x}))^2]$$

There was an error on the slide presented on Tuesday, Sept 20th.

When using this model class, measures how well you expect the "average prediction" to represent the true solution

We expect the bias to decreases with a more complex model

Expected Value

- Linearity of expectation: $E[A + B] = E[A] + E[B]$
- Given a constant c, then $E[cA] = cE[A]$

iid: each example "has the same probability distribution as the others and all are mutually independent."

Generalization Bound for classification

Suppose our test set contained K randomly chosen examples

then by using *Hoeffding's* inequality

the probability our E_{out} differs from E_{test} by more than $\epsilon > 0$ occurs with probability at most $2e^{-2\epsilon^2 K}$

Example:

If K=500 and $\epsilon = 0.1$, then setting $\delta = 2e^{-2(0.1)^2(500)} = 0.0001$ then with probability $1-\delta$ the true error is within 0.1 of the average error on the test set.

Generalization

Cannot get a range - instead get a confidence interval

Hoeffding inequality (stated without proof): for any sample size K, where each random variable is bounded in [a,b] the probability that the average value, v, of the random variables will deviate from its average μ by more than ϵ is:

$$P\big[\,|\,v - \mu\,| > \epsilon\big] \leq 2e^{-2\epsilon^2 K/(b-a)^2} = \delta \text{ for any } \epsilon > 0$$

Thus if $K \geq \dfrac{\log(2/\delta)(b-a)^2}{2\epsilon^2}$ then with probability $1-\delta$

v is ϵ close to μ

We are assuming the K examples are drawn iid from a distribution

Example:
Let g be a binary classifier (g outputs 0,1), let v be the average error of g on the test set of size K, and let μ be the true error of g. The probability that $|v - \mu| > \epsilon$ is at most $2e^{-2\epsilon^2 K}$

If K=500 and $\epsilon = 0.1$, then setting $\delta = 2e^{-2(0.1)^2(500)}$ then with probability $1 - \delta$ the true error is within 0.1 of the average error on the test set.

Generalization

Our estimated average error on our test set

Bound using numbers: K, ϵ and range of output values of function

Cannot get a range - instead get a confidence interval

Hoeffding inequality (stated without proof): for any sample size K, where each random variable is bounded in [0, 1] [a,b] the probability that the average value, v, of the random variables will deviate from its average μ by more than ϵ is:

$$P\left[\,|v-\mu| > \epsilon\right] \le 2e^{-2\epsilon^2 K/(b-a)^2} = \delta \text{ for any } \epsilon > 0$$

Thus if $K \ge \dfrac{\log(2/\delta)(b-a)^2}{2\epsilon^2}$ then with probability $1-\delta$

v is ϵ close to μ

We are assuming the K examples are drawn iid from a distribution

Example:
Let g be a binary classifier (g outputs 0,1), let v be the average error of g on the test set of size K, and let μ be the true error of g. The probability that $|v-\mu| > \epsilon$ is at most $2e^{-2\epsilon^2 K}$

If K=500 and $\epsilon = 0.1$, then setting $\delta = 2e^{-2(0.1)^2(500)}$ then with probability $1-\delta$ the true error is within 0.1 of the average error on the test set.