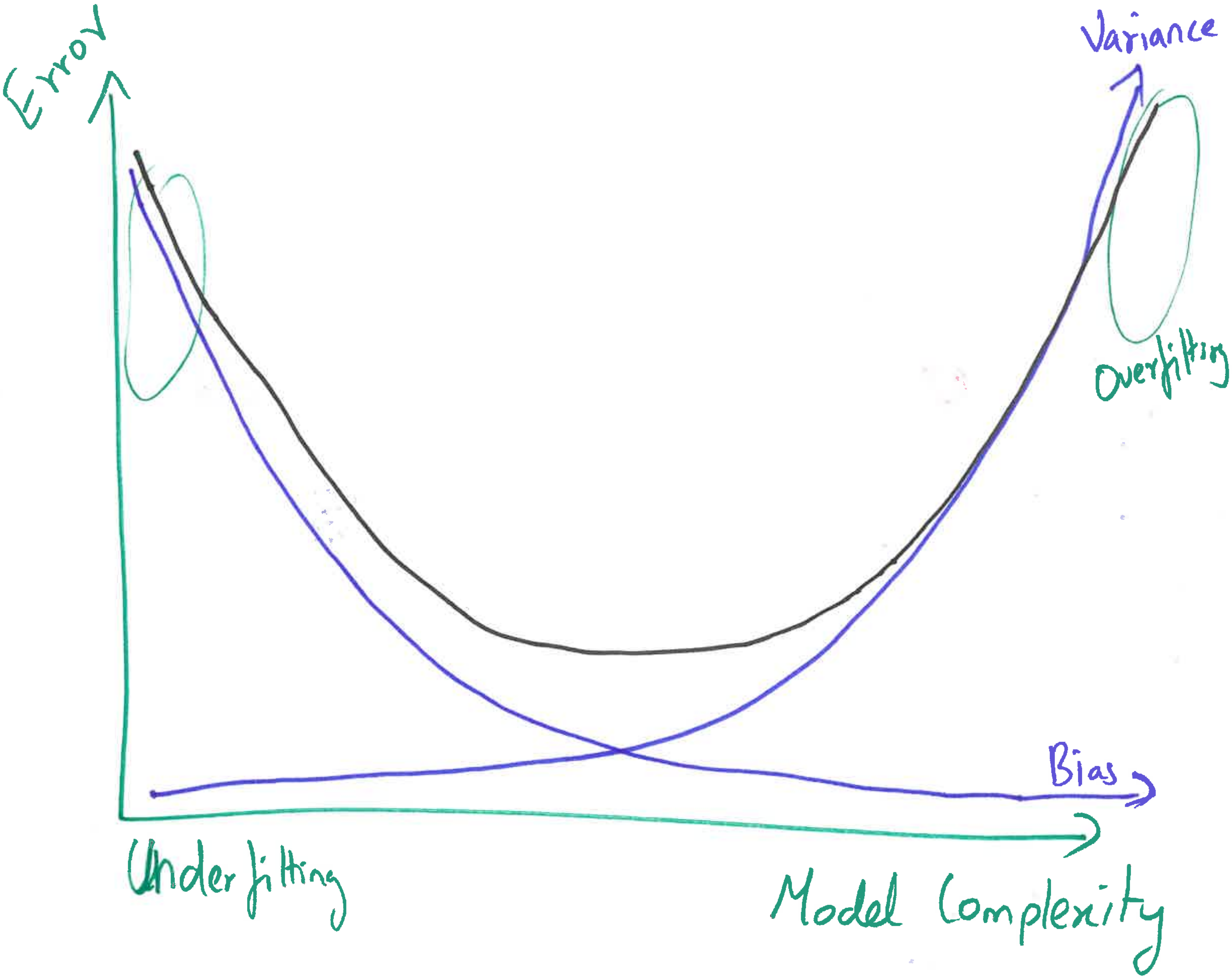
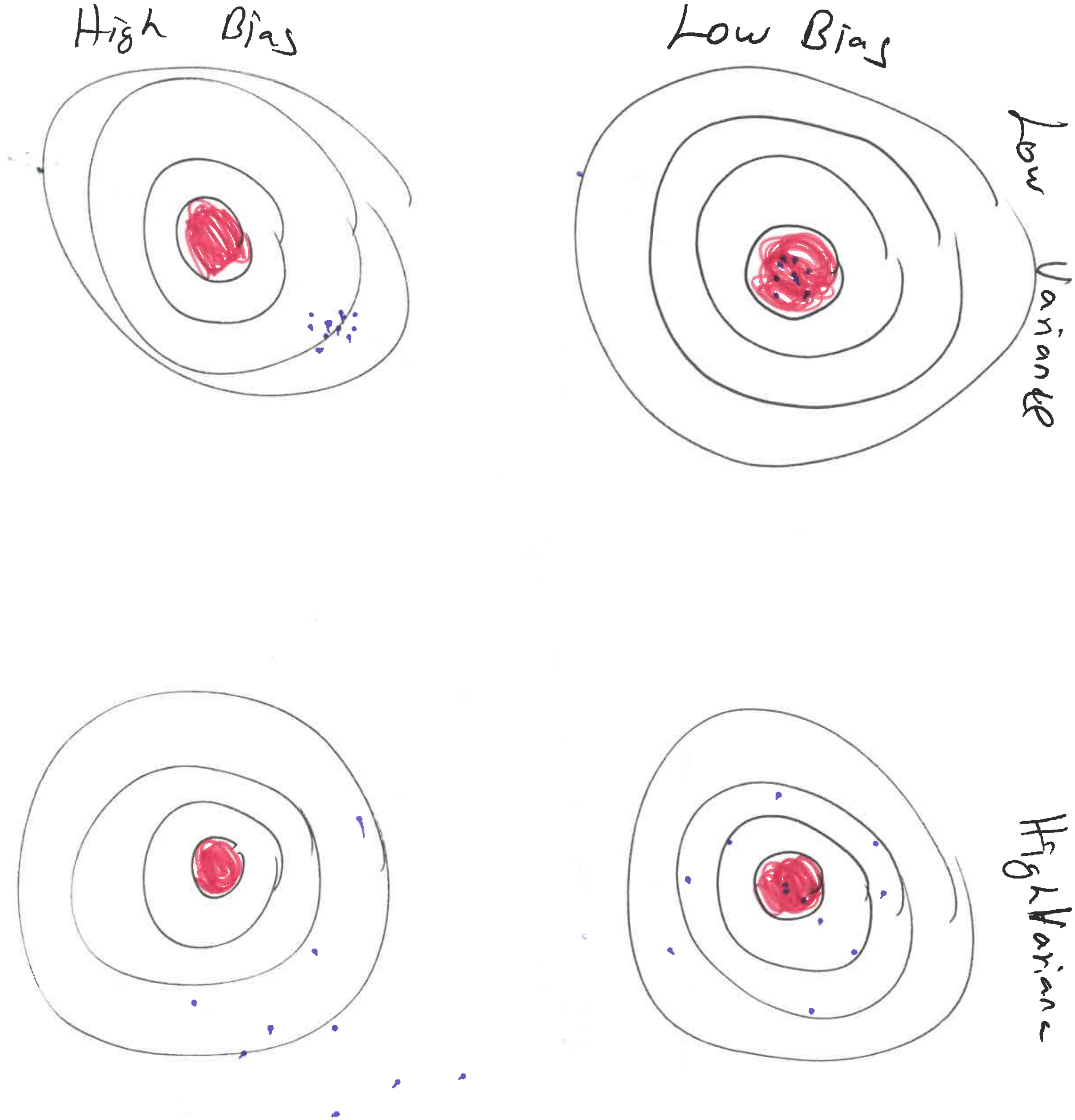
Bias - Variance Trade off




High Bias
Low Bias
Low Variance
High Variance

Two sources of error

① Bias :- wrong assumptions in the Learning algorithm

high Bias → Underfitting

2) Variance : Error from sensitivity to small fluctuation in training set.

High Variance → Overfitting

$$y = f(x) \pm \varepsilon$$

$f(x)$ = true function

$\hat{f}(x)$ = estimate of Y

$$\underline{E\left[\left(f(x) - \hat{f}(x)\right)^2\right]} = \underline{\left(\text{Bias}\left(\hat{f}(x)\right)\right)^2} + \text{Var}\left(\hat{f}(x)\right) +$$

$$\underbrace{\sigma^2} \Rightarrow \text{Irreducible error}$$

$$y = f(x) + \varepsilon$$

$$\varepsilon = \mathcal{N}(0, \sigma^2)$$

Prediction $\hat{f}(x)$

Bias

Suppose S is a statistic which estimates θ → unknown parameter

$$\text{Bias}(S) = \underline{\theta} - E(S)$$

$$\text{Var}(S) = E\left[\left(S - E(S)\right)^2\right]$$

$$\mathrm{Var}(S) = E\left((S - E(S))^2\right)$$

$$\mathrm{Var}(S) = E\left(S^2 - 2SE(S) + (E(S))^2\right)$$

$$\mathrm{Var}(S) = E(S^2) - 2E(S)\,E(S) + (E(S))^2$$

$$\mathrm{Var}(S) = E(S^2) - (E(S))^2$$

$$E(S^2) = \mathrm{Var}(S) + (E(S))^2$$

$$\varepsilon = \mathcal{N}(0, \sigma^2)$$

$$y = f(x) + \varepsilon$$

$$E(y) = f(x)$$

$$\underline{Var(y)} = Var(f(x) + \varepsilon)$$

$$= Var(f(x)) + Var(\varepsilon)$$

$$= 0 + \sigma^2 = \sigma^2$$

Expected value of Squared Error

$$E\left(\left(y-\hat{f}(x)\right)^2\right) = E\left(y^2 - 2\cdot y\cdot \hat{f} + \hat{f}^2\right)$$

$$= E(y^2) - 2\cdot E(y)\cdot E(\hat{f}) + E(\hat{f}^2)$$

$$= \underline{Var(y)} + (E(y))^2 - 2\cdot f\cdot E(\hat{f}) + \underline{Var(\hat{f})} + (E(\hat{f}))^2$$

$$= Var(y) + f^2 - 2\cdot f\cdot E(\hat{f}) + (E(\hat{f}))^2 + Var(\hat{f})$$

$$= Var(\hat{f}) + (f - E(\hat{f}))^2 + Var(y)$$

$$= Var(\hat{f}) + (Bias(\hat{f}))^2 + Var(y)$$

$$E\left((y - \hat{f})^2\right) =$$

$$\underbrace{Var(\hat{f})}_{} + \underbrace{(Bias(\hat{f}))^2}_{} + \underbrace{\sigma^2}_{}$$

$$y = f(x) \pm \varepsilon$$

$$f(x) = x^2$$