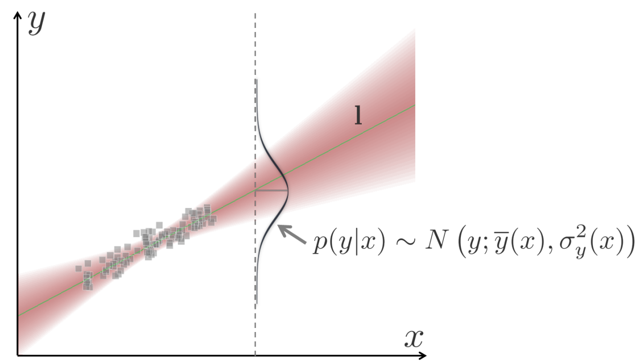

Fig. 4.4: **Probabilistic line fitting.** Given a set of training points we can fit a line $\mathbf{l}$ to them, *e.g.* by least squares or RANSAC. In this example $\mathbf{l} \in \mathbb{R}^2$. Matrix perturbation theory (see appendix A) enables us to estimate a probabilistic model of $\mathbf{l}$ from where we can derive $p(y|x)$ (modelled here as a Gaussian). Training a regression tree involves minimizing the uncertainty of the prediction $p(y|x)$ over the training set. Therefore, the training objective is a function of σ_y^2 evaluated at the training points.

different tasks within the same, general probabilistic forest model. To fully characterize our regression forest model we still need to decide how to split the data arriving at an internal node.

The weak learner model. As usual, the data arriving at a split node j is separated into its left or right children (see fig. 4.1b) according to a binary weak learner stored in an internal node, of the following general form:

$$h(\mathbf{v}, \boldsymbol{\theta}_j) \in \{0, 1\}, \tag{4.4}$$

with 0 indicating "false" (go left) and 1 indicating "true" (go right). Like in classification here we consider three types of weak learners: (i) axis-aligned, ii) oriented hyperplane, (iii) quadratic (see fig. 4.5 for an illustration on 2D$\rightarrow$1D regression). Many additional weak learner models may be considered.

Next, a number of experiments will illustrate how regression forests