1) 5.4 解：设 $error_D(h)$ 用 e 表示，$0.2 \le e \le 0.6$

由公式有 $2 \times 1.96 \times \sqrt{\frac{e(1-e)}{n}} < 0.1$

$\Rightarrow \sqrt{n} > 39.2\sqrt{e(1-e)} \ge 39.2 \times \sqrt{0.2 \times (1-0.2)}$

$\Rightarrow n > 39.2^2 \times 0.2 \times 0.8 \approx 246$

∴ 最少在搜集 246 个样例。

注：本题对题意有另一种理解方式，即需要保证对于 0.2~0.6 范围内的任意 e 均满足要求，这样有不同的答案，即：

$\sqrt{n} > (39.2\sqrt{e(1-e)})_{max}$

而 $39.2\sqrt{e(1-e)} \le 39.2 \times \sqrt{0.5 \times (1-0.5)} = 19.6$

∴ 在这种题意理解下，$n > 19.6^2 \approx 385$，即最少搜集 385 个样例。

2) 答：1. 5% 训练集：$error_D \doteq \frac{2970}{16281} \approx 0.182421$

100% 训练集：$error_D \doteq \frac{2983}{16281} \approx 0.183220$

下分别计算 90%，95% 和 99% 置信区间

	5% 训练集	100% 训练集
90% 置信区间	$0.182421 \pm 1.64 \times \sqrt{\frac{0.182421 \times 0.817579}{16281}} \approx 0.182421 \pm 0.004864$	$0.183220 \pm \sqrt{\frac{0.183220 \times 0.816780}{16281}} \times 1.64 \approx 0.183220 \pm 0.004872$
95% 置信区间	$0.182421 \pm 1.96 \times \sqrt{\frac{0.182421 \times 0.817579}{16281}} \approx 0.182421 \pm 0.005732$	$0.183220 \pm \sqrt{\frac{0.183220 \times 0.816780}{16281}} \times 1.96 \approx 0.183220 \pm 0.005842$
99% 置信区间	$0.182421 \pm 2.58 \times \sqrt{\frac{0.182421 \times 0.817579}{16281}} \approx 0.182421 \pm 0.007550$	$0.183220 \pm \sqrt{\frac{0.183220 \times 0.816780}{16281}} \times 2.58 \approx 0.183220 \pm 0.007822$

2. $\hat{d} = 0.183220 - 0.182421 = 0.000799$

$\sigma_{\hat{d}} \approx \sqrt{\frac{0.182421 \times 0.817579}{16281} + \frac{0.183220 \times 0.816780}{16281}} \approx 0.004284$

$\therefore Z_n = \frac{\hat{d}}{\sigma_{\hat{d}}} = \frac{0.000799}{0.004284} \approx 0.187$

而 $\int_{-0.187}^{0.187} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \approx 0.148$

∴ 双侧区间置信度为 14.8%，单侧区间置信度为 $1 - \frac{100\% - 14.8\%}{2} = 57.4\%$

∴ 5% 训练集性能比 100% 训练集好的可能性为 57.4%。