Задание 2. Поляков Даниил, совместно с 494 гр.

1.

2. **Линейные модели в деревьях** Изначально неизвестно, насколько зависимость в данных близка к линейной. Поэтому, если в листе оказалось много объектов, которые не задают линейную зависимочсть (см. пример из лекции с синусоидой), то линейная регрессия в листьях может не давать ощутимого улучшения. Как модифицировать разбиение по MSE. Как уже говорилось, если в разбиении будет много объектов, то может оказаться так, что линейная зивисимочть плохо их описывает (т.к. линейной зависимости нет вообще или в листе много шумовых объектов). С другой стороны, если в лист попадет слишком мало объектов, то велика вероятность переобучения модели (напр. если в листе будут два объекта, одна из которых шумовая, то лин.регрессия выдаст результат не лучший, чем средний константный ответ.

   Вывод: нужно стоить разбиение выборки так, чтобы число объектов в листе было не слишком малым, но и не слицком большим.

3. Энтропия для непрерывного распределения с плотностью $p(x)$ есть:

   $$H(f) = -\int_{-\infty}^{\infty} p(x)\ln p(x)dx = -\mathbb{E}(\ln(p(x)))$$

   Плотность нормального многомерного распределения есть

   $$p(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)}, \quad x \in \mathbb{R}^n$$

   Тогда

   $$\ln p(x) = \ln\left[\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)}\right] = -\ln\left((2\pi)^{n/2}|\Sigma|^{1/2}\right) - \frac{1}{2}(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)$$

   $$\mathbb{E}(\ln(p(x))) = -(2\pi)^{n/2}|\Sigma|^{1/2} - \frac{1}{2}\mathbb{E}\left[(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)\right]$$

   $$(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu) = \mathbf{x}^\top\Sigma^{-1}\mathbf{x} - 2\mathbf{x}^\top\Sigma^{-1}\mu + \mu^\top\Sigma^{-1}\mu \tag{1}$$

   Используем тот факт, что, если $A-$ симметричная матрица, то $\mathbb{E}(\mathbf{x}^\top A\mathbf{x}) = Tr(A\Sigma) + \mu^\top A\mu$. С учетом этого факта и с учетом выражения (1) получим следующее:

   $$\mathbb{E}\left[(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)\right] = Tr(\Sigma^{-1}\Sigma)+\mu^\top\Sigma^{-1}\mu-2(\mathbf{x}\Sigma^{-1}\mu+\mu^\top\Sigma^{-1}\mu = Tr(I)+2\mu^\top\Sigma^{-1}\mu-2\mu^\top\Sigma^{-1}\mu =$$
   $$= Tr(I) = n$$

   Таким образом

   $$H(f) = -\mathbb{E}(\log(p(x))) = \ln\left((2\pi)^{n/2}|\Sigma|^{1/2}\right) + \frac{n}{2} = \frac{1}{2}\ln((2\pi)^n|\Sigma| + \frac{1}{2}\ln e^n = \frac{1}{2}\ln((2\pi e)^n|\Sigma|)$$