Hamming Code.

Hamming(n, k) with block length $n = 2^m - 1$ and messages length $k = n - m$ is a linear code that maps

$$G : \mathbb{F}_2^k \mapsto \mathbb{F}_2^n \ ,$$

i.e. messages of length k are mapped to codewords of length n. G is injective, so $C = G(\mathbb{F}_2^k)$ is a k–dim. subspace in $\mathbb{F}_2^n$. Usually a systematic code is used where a codeword consists of k message bits together with $m = n - k$ additional parity bits.

It can correct 1 error *or* detect up to 2 errors. The minimum distance between codewords is 3, and for each word $w \in \mathbb{F}_2^n$ there is a codeword $c \in C = G(\mathbb{F}_2^k)$ with distance $d_h(w, c) \leq 1$ (d_h being the Hamming–distance).

G is the *generator matrix*, and there is a *parity check matrix* $H : \mathbb{F}_2^n \mapsto \mathbb{F}_2^m$ such that $Hc = 0 \iff c \in C$. If $w \in \mathbb{F}_2^n$, the vector Hw is called the *syndrome*.

Let c be the transmitted codeword and $w = c + e$ the received word. If there is no error, then $e = 0$. If there is 1 error, i.e. the vector e has only one non-zero entry, then e is equal to the canonical unit vector u_j and $Hw = Hc + He = He = Hu_j$ is equal to the j-th column of the matrix H. If there are more than 1 errors, then w is either another (valid) codeword and $Hw = 0$, or it has distance 1 to another codeword and Hw is also reproduced by different unit error vector, and the decoder will make an error.

So if the received codeword has 2 errors, it will be decoded to the wrong codeword. A parity bit can be added such that the *extended* Hamming code can correct 1 error and detect 2 errors, *or* it can detect up to 3 errors. The distance between codewords is at least 4, so we always have $d_h(w, c) \leq 2$ for some $c \in C$, and if $d_h(w, c) \leq 1$, w can be corrected. If $d_h(w, c) = 2$, a soft decision can be made, if there is additional score/confidence data for the received bits. Then the codeword $c \in C$ with $d_h(c, w) = 2$ can be found which matches best with respect to a metric.

Example: **extended Hamming Code (8,4)**

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad , \quad H = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$C = G(\mathbb{F}_2^4)$ has 16 codewords. Further there are $16 \cdot 8 = 128$ elements in $\mathbb{F}_2^8$ with Hamming–distance $d_h = 1$ to C (i.e. 1-error-words) and $7 \cdot 16 = 112$ elements with $d_h = 2$ (i.e. 2-error-words). If w is a word having 2 errors, then there are 4 codewords c with $d_h(c, w) = 2$.

Soft decision.

For each received bit the decoder produces a score $s_j \in \mathbb{R}$ and makes a hard decision $h_j \in \{-1,+1\}$ (or bit-decision $\hat{h}_j \in \{0,1\} = \mathbb{F}_2$):

$$\begin{array}{lcccl} s_j \geq 0 & \rightsquigarrow & \hat{h}_j = 1 & , & h_j = 2\hat{h}_j - 1 = +1 \\ s_j < 0 & \rightsquigarrow & \hat{h}_j = 0 & , & h_j = 2\hat{h}_j - 1 = -1 \end{array}$$

Instead of using the algebraic properties of the linear code and the Hamming–distance to decoded the received bit-word $\hat{h} = (\hat{h}_1, \dots, \hat{h}_n)$, one can also use the bit-scores of a soft decision decoder and a different distance function to find the best matching codeword by consider the codewords in $\{-1,+1\}^n \subset \mathbb{R}^n$ and the scores of the received word in $\mathbb{R}^n$.
Let $s = (s_1, \dots, s_n)$ be the received *soft* word and $h = (h_1, \dots, h_n) \in \{-1,+1\}^n$ with $s_j = h_j|s_j|$ the corresponding *hard* word:

$$h_j = \frac{s_j}{|s_j|} = \operatorname{sgn} s_j \qquad (s_j \neq 0)$$

$$|h_j| = 1 = h_j h_j \quad , \quad s_j = h_j|s_j| \quad \rightsquigarrow \quad |s_j| = h_j s_j$$

If $y = (y_1, \dots, y_n) \in \{-1,+1\}^n$ is another hard word, then $\operatorname{corr}(s,y) \leq \operatorname{corr}(s,h)$, where

$$\operatorname{corr}(s,h) = \sum_j s_j h_j = \sum_j |s_j| = \|s\|_1 \geq 0 \ .$$

The best *valid* match $y \in \{-1,+1\}^n$ *maximizes* $\operatorname{corr}(s,y)$. Using $y_j = \pm h_j$, we have

$$\operatorname{corr}(s,h) - \operatorname{corr}(s,y) = \sum_j s_j(h_j - y_j) = \sum_{h_j \neq y_j} s_j(h_j - y_j) = 2\sum_{h_j \neq y_j} |s_j| \geq 0 \ .$$

Thus the best match y is for which the sum of $|s_j|$ is minimal for $y_j \neq h_j$, i.e. the errors are probably at positions with lower scores:

$$\operatorname{corr}(s,y) = \max_{\hat{x} \in C}\{\operatorname{corr}(s,x)\} \leq \operatorname{corr}(s,h) \ .$$

It is also possible to use the Euclidean distance d_2 or Manhattan distance d_1, though for d_1 the scores s_j need to be normalized.
For $h, y \in \{-1,+1\}^n$ and $s_j = h_j|s_j|$ we have

$$\begin{aligned} d_p(s,h)^p &= \sum_j |s_j - h_j|^p = \sum_j \big||s_j|h_j - h_j\big|^p = \sum_j \big||s_j| - 1\big|^p \ , \\ d_p(s,y)^p &= \sum_j |s_j - y_j|^p = \sum_j \big||s_j|h_j - y_j\big|^p \\ &= \sum_{h_j = y_j} \big||s_j|h_j - h_j\big|^p + \sum_{h_j \neq y_j} \big||s_j|h_j + h_j\big|^p \\ &= \sum_{h_j = y_j} \big||s_j| - 1\big|^p + \sum_{h_j \neq y_j} \big||s_j| + 1\big|^p \\ &\geq d_p(s,h)^p \quad . \end{aligned}$$

Since

$$d_1(s,y) - d_1(s,h) = \sum_{h_j \neq y_j} \big(\big|1 + |s_j|\big| - \big|1 - |s_j|\big| \big) \geq 0 \ ,$$

a soft decision for d_1 is only possible for $s_j \in [-1, +1]$, i.e. if the bit-scores are normalized. Then choose *valid* $\hat{y} \in C$ such that $d_1(s,y)$ is *minimal*,

$$d_1(s,y) = \min_{\hat{x} \in C} \{ d_1(s,x) \} \ .$$

For d_2 we get

$$\begin{aligned} d_2(s,y)^2 - d_2(s,h)^2 &= \sum_{h_j \neq y_j} \big((1 + |s_j|)^2 - (1 - |s_j|)^2 \big) \\ &= \sum_{h_j \neq y_j} \big(2|s_j| + 2|s_j| \big) = 4 \sum_{h_j \neq y_j} |s_j| \geq 0 \ , \end{aligned}$$

which leads to the same soft decision as $\operatorname{corr}(s,h) - \operatorname{corr}(s,y)$ for $s \in \mathbb{R}^n$,

$$d_2(s,y)^2 - d_2(s,h)^2 = 2\big(\operatorname{corr}(s,h) - \operatorname{corr}(s,y) \big) \ .$$

Choose *valid* $\hat{y} \in C$ such that $d_2(s,y)$ is *minimal*,

$$d_2(s,y) = \min_{\hat{x} \in C} \{ d_2(s,x) \} \ .$$

If soft decision is frequently used for 2-error words, then it is likely that 3 errors occur that will be decoded to the wrong codeword. Thus for higher error rates, e.g. an additional CRC over several codewords can give a second opinion.